

**Register awareness and English
language learning: the case of
multi-word discourse markers**

Anne Li-E Liu

**Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy**

May, 2013

Dedicated to the memory of Rai-Yun Zeng

(March 1, 1950 – January 9, 2007)

Abstract

Research in L2 learners' use of discourse markers (DMs), which largely focuses on single-word DMs and reports learners' overuse or underuse tendency, generally concludes that L2 learners need a keener register awareness in this regard. This lack of awareness in using DMs in accordance with the appropriate register, however, is not further pursued. Although extensive studies have been carried out in examining multi-word discourse markers (MDMs), researchers have exclusively focused on the formulaic nature of these MDMs. The linking nature that MDMs possess has been largely neglected. This thesis therefore aims to explore further the pragmatic awareness of L2 learners in their use of MDMs by including both corpus-based studies and experiments. Questions to be addressed include to what extent L2 learners exert their pragmatic awareness and use MDMs appropriately, and whether or not English proficiency affects the types and quantities of MDMs used by learners. The thesis first describes the usage patterns of the targeted MDMs in 4 native speaker (NS) corpora, leading to the creation of a formality continuum along which various MDMs can be placed. An additional investigation in the Cambridge Learner Corpus (CLC) shows that the overuse/underuse phenomenon has to be discussed by taking into account the semantic categories of MDMs. Two studies—the reaction time (RT) task and the multiple discourse completion task (MDCT)—are carried out with the goals of perceiving learners' pragmatic awareness and testing whether the learners' corpus-based results can be supported by the experimental data. The results show that advanced L2 learners notice the embedded formality mismatch better than the intermediate learners. The L2 participants in the two studies find identifying Concessive links the most difficult and this is supported by the fact that CLC learners also show a lack of register awareness in using Concessive MDMs. Future work includes applying the methodology to other multi-word units, recruiting different groups of L2 learners (ESL), and pursuing the thesis's implication for teaching.

Acknowledgement

Completing this PhD thesis feels as if I have just given birth to a second baby even though it took three and half years. I could not make it this far without many people's support. The most sincere gratitude goes to my advisor, Professor Svenja Adolphs. She has never hesitated in giving me guidance, support, as well as pointing out directions that I have not thought of. She believed in me at the outset and her non-stop encouragement has guided me through this long journey, through the good and bad times. I particularly appreciate the freedom and trust she has given me in pursuing independent works. I owe her my deepest gratitude. In addition, special thanks to my committee, Professor Ronald Carter and Dr Ivor Timmis for their encouragement, guidance and helpful suggestions. Specifically, Prof Carter is another important figure in this journey. I thank him for his insightful comments that he has provided as an examiner for my annual review, and his continued support since then. I also want to express my gratitude to Professor Norbert Schmitt for his encouragement, guidance, and support.

My friends in the UK, Taiwan, and other parts of the world have also made their contribution to this journey in various ways. I thank Dr Nai-Lung Tsai for his knowledge and assistance in programming. I also want to express my gratitude to Professor David Wible, my MA thesis advisor, for his friendship and confidence in me. Special thanks to Nathalie Dubois, Paweł Szudarski, Wenjuan Yuan, Eric Yen-liang Lin, Yung-Wen Yao, Klaudia Lee, Daniel Hunt, Giulia Privitera, Ania Kaminska, Letty Chen, and Jib Wipapan for the laughter, joy, and support you have provided.

Last but not the least thanks goes to my family in Taiwan and my beloved husband, Jonathan Good. Without your undivided love and faith in me, I would not have made it this far.

Contents

List of Figures	x
List of Tables	xi
Chapter 1 – Introduction.....	1
1.1 Setting the Scene.....	1
1.2 Aims of the thesis.....	9
1.3 Structure of the thesis	11
Chapter 2 - Discourse Markers and Interlanguage Pragmatics.....	13
2.1 What are discourse markers?.....	13
2.2 Features of DMs.....	15
2.2.1 Connectivity.....	15
2.2.2 Optionality	16
2.2.3 Orality	17
2.2.4 Multi-categoriality	17
2.3 Two main accounts on DMs	19
2.4 DMs in context—a pragmatic perspective.....	22
2.5 Corpus-based studies of L2 learners' use of DMs.....	25
2.6 Pragmatic competence	31
2.6.1 Hyme's communicative competence	33
2.6.2 Canale and Swain's communicative competence.....	34
2.6.3 Bachman's communicative competence	35
2.6.4 Celce-Murcia, Dörnyei, and Thurrell's communicative competence	37
2.7 Interlanguage pragmatics	40
2.7.1 Methodologies in ILP	41
2.7.2 Factors in developing interlanguage pragmatics.....	42
2.8 The adopted theoretical frameworks	43
2.8.1 Noticing hypothesis.....	43
2.8.2 Accommodation theory	44
2.9 Summary and conclusion	46

Chapter 3 - Towards a Description of Multi-word Discourse Markers in Context	48
3.1 MDMs and registers: situated language use of MDMs	49
3.2 Altenberg's formality scale	52
3.3 The four corpora	55
3.3.1 The Cambridge and Nottingham Corpus of Discourse in English (CANCODE)	56
3.3.2 The academic discourse in British National Corpus (BNC-Acad)	56
3.3.3 The British Academic Spoken English (BASE) Corpus	57
3.3.4 The Enronsent2 corpus	57
3.4 Operational definition of multi-word discourse markers (MDMs)	59
3.5 The target MDMs	60
3.6 Methods of data analysis	62
3.6.1 Qualitative approach	62
3.6.2 Quantitative approach	62
3.6.2.1 Frequency cut-off point and the screening process	62
3.6.2.2 Additive smoothing—the <i>add one</i> rule used for zero frequency	64
3.6.2.3 Methods in calculating formality scores	65
DIFFERENCE COEFFICIENT (<i>diff. coeff.</i>)	66
BROOKE, WANG, AND HIRST'S SIMPLE FORMALITY SCORE (<i>FS</i>)	66
WEIGHTED FORMALITY AVERAGE METHOD (<i>WFA</i>)	67
3.7 Results and discussion	71
3.7.1 Additive MDMs	71
3.7.2 Appositive MDMs	74
3.7.3 Causal MDMs	76
3.7.4 Concessive MDMs	78
3.7.5 Contrastive MDMs	80
3.8 The comparisons of <i>WFA</i> , <i>diff. coeff.</i> , and <i>FS</i> methods	82
3.9 The formality continuum	85
3.10 Summary and conclusion	87
Chapter 4 - Discourse markers in use: evidence from Cambridge Learner Corpus	89
4.1 The Cambridge Learner Corpus (CLC)	93

4.2 CLC data in the formal and the informal register	96
4.2.1 Additive MDMs.....	96
4.2.1.1 Formal register.....	96
4.2.1.2 Informal register.....	101
4.2.2 Appositive MDMs	104
4.2.2.1 Formal register.....	104
4.2.2.2 Informal register.....	107
4.2.3 Causal MDMs.....	109
4.2.3.1 Formal register.....	109
4.2.3.2 Informal register.....	112
4.2.4 Concessive MDMs.....	115
4.2.4.1 Formal register.....	115
4.2.4.2 Informal register.....	121
4.2.5 Contrastive MDMs	125
4.2.5.1 Formal register.....	125
4.2.5.2 Informal register.....	128
4.2.6 Putting corpus data together	130
4.2.7 summary.....	136
4.3 Comparing the learner data with NS data.....	138
4.3.1 Additive MDMs.....	140
4.3.2 Appositive MDMs	143
4.3.3 Causal MDMs.....	146
4.3.4 Concessive MDMs.....	149
4.3.5 Contrastive MDMs	151
4.4 Discussion.....	154
4.4.1 The frequencies of MDMs in the five semantic categories	154
4.4.2 The formality makeup of the MDMs	159
4.5 Summary and conclusion.....	163
Chapter 5 - Reflecting L2 pragmatic awareness in two registers—an experimental approach using corpus-informed material.....	166

5.1 Theoretical framework	170
5.1.1 Noticing hypothesis.....	170
5.1.2 Accommodation theory	171
5.2 The two registers.....	173
5.3 Study 1: a reaction time design	174
5.3.1 Participants.....	175
5.3.2 Materials	175
5.3.3 Procedure	179
5.3.4 Measures.....	179
5.3.5 A common ground for comparing spoken with written language.....	180
5.3.6 Results and Discussion	184
5.3.6.1 The mean RT in the 12 conditions	184
5.3.6.2 Mean RT in formality-consistent conditions	187
5.3.6.3 Mean RT in formality-inconsistent conditions	188
5.3.6.4 Paired samples T-tests	189
<i>NS data</i>	190
<i>L2 data</i>	191
<i>Discussion</i>	192
5.3.6.5 Two-way ANOVA	195
<i>NS data</i>	195
<i>L2 data</i>	196
5.3.7 Discussion	196
5.3.8 Conclusion.....	198
5.4 Study 2: A multiple-choice discourse completion task (MDCT).....	199
5.4.1 Participants.....	200
5.4.2 Materials	201
5.4.3 Procedure	204
5.4.4 Measures	205
5.4.5 Results.....	207
5.4.5.1 Three-way ANOVA in CLASS measure.....	207

5.4.5.2 Three-way ANOVA in EXACT MATCH measure	209
5.4.5.3 Discussion of the statistical analysis.....	212
5.4.5.4 Implication of the mean scores in the CLASS measure	213
5.4.5.5 Difference between the CLASS measure and the EXACT MATCH measure	214
5.4.6 Discussion	219
5.5 Putting Studies 1 and 2 together.....	221
5.6 Summary and conclusion	224
Chapter 6: Putting corpus evidence and experiments together	226
6.1 Study 1 and the corpus evidence	227
6.1.1 A quick recap of Study 1.....	227
6.1.2 Interpreting the meaning of a long and short RT.....	228
6.1.3 Formality-consistent conditions	230
6.1.4 Formality-inconsistent conditions	233
6.1.5 The RT differences between the two types of conditions	236
6.2 Study 2 and the corpus evidence	239
6.2.1 A brief recap of Study 2	239
6.2.2 A macro view of the two measures and the corpus data	241
6.2.3 Interpreting the relationship between the two measures.....	243
6.2.4 The two-measure differences in the formal register and the corpus data.....	245
6.2.5 The two-measures difference in the informal register and the corpus data	248
6.2.6 Discussion	250
6.3 The evaluation of the methodologies in the two studies.....	253
6.4 Summary and conclusion	254
Chapter 7 - Conclusion	257
7.1 A summary of the thesis	257
7.2 General conclusions.....	261
7.3 Limitations.....	263
7.4 Directions for further research and its application	265
7.4.1 Methodology	265
7.4.2 Learner corpus data.....	267

7.4.3 Types of learner.....	267
7.5 Final remarks	269
References.....	270
Appendix 1 List of target MDMs	296
Appendix 2 Study 1: the nine passages for the six target MDMs and the neutral MDMs	297
Appendix 3 Study 1: Experimental list 1	305
Appendix 4 Study 2: Multiple-choice Discourse Completion Task.....	318

List of Figures

Figure 2.1. Communicative competence constructed in Canale and Swain (1980) and Canale (1983)	35
Figure 2.2 Bachman's framework of communicative competence.....	36
Figure 2.3 Celce-Murcia et al.'s schematic representation of communicative competence.....	38
Figure 3.1 The interaction of language use and the registers.....	53
Figure 3.2 The formality continuum of MDMs	86
Figure 4.1 Percentages of formal and informal Additive MDMs, formal register	100
Figure 4.2 Percentages of formal and informal Additive MDMs, informal register	103
Figure 4.3 Percentages of formal and informal Appositive MDMs, formal register.....	107
Figure 4.4 Percentages of formal and informal Causal MDMs, formal register	111
Figure 4.5 Percentages of formal and informal Causal MDMs, informal register.	114
Figure 4.6 Percentages of formal, informal, and neutral Concessive MDMs, formal register.....	118
Figure 4.7 Percentages of formal, informal, and neutral Concessive MDMs, informal register.....	124
Figure 4.8 Percentages of formal and neutral Contrastive MDMs, formal register.....	127
Figure 4.9 Percentages of formal and neutral Contrastive MDMs, informal register.....	129
Figure 4.10 Breakdown of MDMs by formality category, formal register	133
Figure 4.11 Breakdown of MDMs by formality category, informal register	133
Figure 4.12 Formal-informal ratios of Additive MDMs	142
Figure 4.13 Formal-neutral-informal ratios of Appositive MDMs.....	145
Figure 4.14 Formal-informal ratios of Causal MDMs	147
Figure 4.15 Formal-neutral-informal ratios of Concessive MDMs.....	150
Figure 4.16 Formal-neutral ratios of Contrastive MDMs	153
Figure 4.17 Formality makeup of the MDMs produced	161
Figure 5.1 Mean reaction time (RT) per syllable in consistent cases	187
Figure 5.2 NS and L2 participants' mean reaction time (RT) per syllable in inconsistent cases	188

List of Tables

Table 1.1	Examples of conjuncts	5
Table 1.2	Linking words and phrases.....	6
Table 3.1	Stylistic differentiation of Contrastive DMs proposed in Altenberg's study (1986, p. 18)	54
Table 3.2	Summary of the four corpora included in this chapter	56
Table 3.3	Target MDMs shown in the semantic categories (The symbol * refers to MDMs that are not from the four reference books)	61
Table 3.4	The cross-corpus comparisons and the dimension difference factors.....	69
Table 3.5	Results of cross-corpus comparisons of what's more.....	70
Table 3.6	Raw and normalized frequencies of Additive MDMs	72
Table 3.7	Formality scores of Additive MDMs.....	72
Table 3.8	Raw and normalized frequencies of Appositive MDMs.....	75
Table 3.9	Formality scores of Appositive MDMs	75
Table 3.10	Raw and normalized frequencies of Causal MDMs	77
Table 3.11	Formality scores of Causal MDMs.....	77
Table 3.12	Raw and normalized frequencies of Concessive MDMs	78
Table 3.13	Formality scores of Concessive MDMs.....	70
Table 3.14	Raw and normalized frequencies of Contrastive MDMs.....	81
Table 3.15	Formality scores of Contrastive MDMs	81
Table 4.1	The six levels in CEFR.....	94
Table 4.2	Summaries of the seven sub-corpora in CLC	95
Table 4.3	RF and NF of Additive MDMs in CLC, formal register.....	97
Table 4.4	LL test results of Additive MDMs, formal register	98
Table 4.5	RF and NF of Additive MDMs in CLC, informal register	101
Table 4.6	LL test results of Additive MDMs, informal register.....	102
Table 4.7	RF and NF of Appositive MDMs in the four learner corpora, formal register.....	105
Table 4.8	LL test results of Appositive MDMs, formal register	106
Table 4.9	RF and NF of Appositive MDMs in the three learner sub-corpora, informal register.....	107
Table 4.10	LL test results of Appositive MDMs, informal register	108
Table 4.11	RF and NF of Causal MDMs in the four learner corpora, formal register	109
Table 4.12	LL test results of Causal MDMs, formal register	111
Table 4.13	RF and NF of Causal MDMs in the four learner corpora, informal register	112

Table 4.14	LL test results of Causal MDMs, informal register.....	113
Table 4.15	RF and NF of Concessive MDMs in the four learner corpora, formal register.....	116
Table 4.16	LL test results of Concessive MDMs, formal register.....	117
Table 4.17	Use of despite and in spite of as DMs and non-DMs in the four sub-corpora	120
Table 4.18	RF and NF of Concessive MDMs, informal register	122
Table 4.19	LL test results of Concessive MDMs, informal register.....	123
Table 4.20	RF and NF of Contrastive MDMs in the four learner corpora, formal register.....	126
Table 4.21	LL test results of the Contrastive MDMs	126
Table 4.22	RF and NF of Concessive MDMs, informal register	128
Table 4.23	LL test results of Contrastive MDMs, informal register.....	129
Table 4.24	Summary of the total MDMs in each semantic category, formal register	131
Table 4.25	Summary of the total MDMs in each semantic category, informal register.....	132
Table 4.26	Summary of the corpora to be compared.....	139
Table 4.27	NF of Additive MDMs in the NS and learner corpora, F=formal, Neu=Neutral, Inf=informal.....	140
Table 4.28	NF of Appositive MDMs in the NS and learner corpora, F=formal, Neu=Neutral, Inf=informal.....	144
Table 4.29	NF of Causal MDMs in the NS and learner corpora, F=formal, Inf=informal.....	146
Table 4.30	NF of Concessive MDMs in the NS and learner corpora, F=formal, Neu=Neutral, Inf=informal.....	149
Table 4.31	NF of Contrastive MDMs in the NS and learner corpora, F=formal, Neu=Neutral	152
Table 4.32	A summary of all the MDMs in each semantic category, NS data	155
Table 4.33	A summary of all the MDMs in each semantic category, learner data	155
Table 4.34	NF of MDMs by formality.....	159
Table 4.35	The breakdown of learners' MDMs by formality and semantics.....	162
Table 5.1	A summary of the six critical items	176
Table 5.2	The six experimental list (Neu: neutral, F: formal, Inf: informal. FC: formal context, InfC: informal context)	178
Table 5.3	Result of Lexile analysis of the 12 conditions (FM-FR: formal MDM in formal register, FM-InfR: formal MDM in informal register, InfM-FR: Informal MDM in formal register, InfM-InfR: informal MDM in informal register).....	183

Table 5.4	Summary of the NS and L2 participants' mean RT in the 12 conditions	185
Table 5.5	Summary of RT differences in each paired condition	186
Table 5.6	Overview of the pairs compared (F: formal, M: marker, Inf: informal, R: register).....	189
Table 5.7	Descriptive statistics of the paired samples T-test in NNS data	191
Table 5.8	The ten MDMs used in the MDCT	202
Table 5.9	Item order in the MDCT.....	204
Table 5.10	Mean and standard deviation of the 10 conditions in CLASS measure, F: formal, Inf: informal	208
Table 5.11	Mean and standard deviation of the 10 conditions in EXACT MATCH measure, F: formal, Inf: informal	210
Table 5.12	Total mean scores of the five semantic types	213
Table 5.13	Mean score in the informal Causal condition.....	215
Table 5.14	Mean score differences of the two measures in the 5 formal conditions	217
Table 5.15	Mean score differences of the two measures in the 5 informal conditions	218
Table 6.1	RT of the 6 formality-consistent conditions. F: formal, Inf: informal, C: context	231
Table 6.2	RT of the 6 formality-inconsistent conditions. F: formal, Inf: informal, C: context	233
Table 6.3	L2 participants' RT differences between the two conditions. F: formal, Inf: informal, C: context	237
Table 6.4	Total mean scores in the two registers by the two measures	241
Table 6.5	Summary of the L2 participants' mean scores	242
Table 6.6	L2 participants' two-measure differences in the formal conditions	245
Table 6.7	L2 participants' two-measure differences in the informal conditions	248

Chapter 1 - Introduction

1.1 Setting the Scene

When one of my classmates in the MPhil program at the University of Cambridge used the word *hence* in the middle of a casual conversation, I felt a sense of awkwardness. I only use *hence* when writing academic papers that require a formal style. I was told later by friends whose first language is English that the Causal marker *hence* is only used in daily conversation when one intends to be humorous, and therefore the use of *hence* is meant to be entertaining. When expressing a Causal relationship by including a discourse marker in conversation, language users frequently employ markers like *so*, *because of that*, or *for that reason*. I did not ask that classmate his reason for using the marker *hence*, but I am sure that he was not being humorous. In fact, he continued taking part in the conversation without noticing the awkwardness that others felt due to the out-of-place word. This classmate scored 8.5 in IELTS and is considered a very advanced learner of English. Why, then, the lack of awareness as to the proper context for *hence*?

It would require an exhaustive list to consider all the possible challenges that a second language (L2) learner must meet before claiming to be confident in English. Indeed, abundant research has documented various difficulties that learners face, ranging from word-level problems, such as pronunciation and vocabulary size, to structure-level considerations, like syntax and grammar, to discourse-level issues such as coherence and cohesion, discourse strategies, and whether or not the language produced is appropriate in a given context. As English proficiency develops, the last

issue - the notion of *appropriateness in context* (Hymes, 1972; Campbell & Wales, 1970; Dewaele, 2008) - becomes more salient. The anecdote about my former classmate's use of *hence* is a good example to show the predicament of advanced learners who speak and write fluently, but who seem to neglect the pragmatic considerations in discourse.

The evaluation of the language produced by learners with regard to pragmatics is termed Interlanguage Pragmatics (ILP). Employing appropriate words or phrases in accordance with the context has become a central issue in Second Language Acquisition (SLA), especially in the research of pragmatic competence. Many of the studies on pragmatic competence in SLA tend to focus on examining learners' illocutionary competence, i.e., various speech acts that learners perform (compliments: Rose & Ng, 2001; Lorenzo-Dus, 2001; requests: Hassall, 1997; Li, 2000; Rose, 2000; Ellis, 1992; refusals: Félix-Brasdefer, 2004; apologies: Trosborg, 1995; Schauer, 2006). Literature to address learners' sociolinguistic competence (Bachman, 1990), particularly focusing on learners' awareness of register or formality differences, is, to the author's knowledge, very limited. The present thesis deals with this less-explored aspect: learners' sociolinguistic competence. Specifically, I will investigate learners' pragmatic awareness with regard to their use of multi-word discourse markers (MDMs) in the formal and informal registers.

MDMs are multi-word units (*as a result, having said that, on the contrary*) that connect sentences, and at the same time, maintain the flow of discourse. Because of the linking function, discourse markers (DMs) are also called connectives or linking adverbials (Biber et al., 1999). As MDMs also share the same quality that is particularly

attached to formulaic language, "sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated" (Wray, 2002, p. 9), these multi-word units are also referred to as conventional or formulaic expressions.

Although research on how learners use MDMs as connectives or formulaic language is not uncommon, I observe two problems. Research on formulaic expressions usually includes MDMs, such as *on the other hand* and *on top of that*, focusing on whether or not learners process or retrieve these multi-word units as a whole. The connective function that these MDMs have is inevitably neglected. Reviewing studies in DMs shows that the unit of analysis tends to be single-word DMs (SDMs), such as *but* in Birner (1988), *however* and *still* in Lenk (1998), *nevertheless* and *but* in Blakemore (2000), *anyway* in Urgelles-Coll (2010), *so* and *well* in Muller (2005), *nevertheless*, *but*, and *yet* in Bell (2010), and *so* in Bolden (2009), to name just a few. Some notable exceptions are *on the contrary* in Fraser (2009), and *on the other hand* in Bell (2004). The fact that MDMs receive less attention in DM studies, however, should not be interpreted as indicating less importance. Further, even though most of the MDMs are frozen phrases and do not have open slots for variation, one should not have the impression that MDMs are easy to acquire. On the contrary, I argue that the dual roles of MDMs, as formulaic expressions and connectives, make them a central issue in Second Language (L2) studies. The appropriate use of formulaic expressions and DMs is indeed considered as native-like performance (Wray, 1999, 2002, 2004; Wood, 2000; Hyland & Tse, 2004); yet, as in the case mentioned above, DM studies in L2 learners tend to look at SDMs (Bikeliene, 2008; Tang & Ng, 1995; Field & Yip, 1992; Milton & Tsang, 1993; Altenberg & Tapper, 1998; Granger & Tyson, 1996). There is

clearly a dearth of description of MDMs in general, which in turn highlights the need for research into this area. This thesis aims to provide a thorough description of MDMs in use, and further, to explore L2 learners' use of various MDMs as well as how learners show their pragmatic competence via their use of MDMs.

Sinclair discussed the reason why speakers and writers choose one lexical unit over another, attributing such choice to semantic prosody, which is "a subtle element of attitudinal, often pragmatic meaning" (2004, p. 145). Semantic prosody, in his accounts, expresses the function of a lexical item, and is on the "pragmatic side of the semantic/pragmatic continuum" (Sinclair, 2004, p. 34). He exemplified the above argument by proposing that a language speaker who uses the verb *budge* rather than *move* shows his attitude, and such a choice reflects the speaker's pragmatic consideration. Thinking along the same line as Sinclair, I propose that one's choice of MDM also shows the speaker's consideration in semantics and pragmatics. While the semantic factor decides whether a speaker, for example, chooses a Causal, a Contrastive, or a Concessive marker, the pragmatic factor refers to how a speaker takes into account the formality of a context and makes an accordingly appropriate choice. I will use corpus data to illustrate this.

The data obtained from the CANCODE (Cambridge and Nottingham Corpus of Discourse English) corpus, which is used in the present thesis, shows that *on top of that* and *what's more* are Additive MDMs employed rather frequently in spoken discourse, but both occur with a relatively low frequency in academic written discourse (data from the academic section of the British National Corpus). When writers are to use an Additive MDM, they prefer *in addition*. In other words, while the

semantic consideration indicates that a language user should employ an Additive MDM, it is the pragmatic factor that decides which Additive MDM to use. Should the context be formal, a language user will be more likely to choose *in addition* rather than *on top of that*.

Such usage difference in registers might be an intuitive one for native speakers of English (NS); this subtle difference in formality, however, is not commonly pinpointed for L2 learners. In an EFL or ESL curriculum, these markers are normally introduced to learners as transitional words and phrases, or as stylistic devices (Reid, 1993; Hinkel, 2001). Markers that are classified under the same semantic category are conventionally presented to learners as a group in dictionaries or grammar books. Additional information with regard to their situated uses is exemplified with only a single sample sentence, if at all. An example of such presentation is given by Greenbaum (1996), who introduced various conjuncts (DMs in this study) by listing them on the basis of their meaning (Table 1.1).

Table 1.1

Examples of conjuncts (Greenbaum, 1996, p. 146)

first, second, ...firstly, secondly...; next, then, finally..
equally, likewise, similarly, in the same way
again, also, further, furthermore, moreover, what is more, in addition
so, therefore, thus; hence, consequently, as a result, as a consequence, in consequence

Even though Greenbaum did not describe the semantic categories for each, the four columns refer to Enumerative, Appositive, Additive, and Causal markers respectively. Swales and Feak (2004) presented examples of linking devices in a similar manner, shown in Table 1.2.

Table 1.2
Linking words and phrases (Swales & Feak, 2004, p. 22)

Sentence connectors	
<i>Addition</i>	furthermore, in addition, moreover
<i>Adversative</i>	however, nevertheless
<i>Cause and effect</i>	therefore, as a result, consequently, hence, thus
<i>Clarification</i>	in other words, that is, i.e.

Although these markers/connectors are labelled according to the semantic classification, they are presented to the readers, and possibly to L2 learners, without any further information that differentiates one from another. For example, in Greenbaum's list, the less formal *so* is juxtaposed with more formal *hence* and *in consequence*. This list/table presentation runs the risk of giving a misleading impression that these markers form a homogeneous group, and are interchangeable to some extent. Zamel (1983) and Crewe (1990) pointed out that the misleading presentation in textbooks only leads to more problematic uses of DMs in L2 writing. This position is supported by the results of studies that examine DMs used by L2 writers. A consistent conclusion that emerges from these studies is that L2 learners

show a tendency to use informal DMs in argumentative essays (Field & Yip, 1992; Granger & Tyson, 1996; Altenberg & Tapper, 1998; Milton & Tsang, 1993; Bikeliene, 2008), indicating a lack of register awareness in learners' use of DMs.

Most of the aforementioned studies that investigate DMs used by L2 learners, however, tend to make inferences on the basis of the comparisons made between NS writing and L2 writing, and report an overuse/underuse dichotomy. The register-related pragmatic awareness in using markers is not pursued further. In addition, the conclusion that L2 learners use informal DMs in writing is drawn based on SDM data. Whether or not L2 learners also show a lack of register awareness in using MDMs remains to be answered. This study aims to examine this under-explored area, and to provide a better understanding and description of how learners use MDMs from a cognitive perspective.

Conrad (2004) examined the use of *though* on the basis of Biber et al.'s (1999) corpus evidence and four ESL textbooks. According to Biber et al. (1999), *though* is the most common Concessive conjunction in speech (based on the 6.5 million words of English conversation), whereas it is *however* that is the most frequently occurring marker in academic prose. Even though the four ESL textbooks Conrad selected claim to be helpful for learners in both spoken and written language, Conrad discovered that only one textbook introduces the contrastive use of *though*, and none of them mentions its Concessive use. The difference observed between the corpus evidence and textbook lessons further illustrates the previous point that L2 learners receive limited, and sometimes misleading, input, which is likely to lead to L2 learners' inappropriate use of markers.

Determining the formality of or register preference in SDMs can be carried out by taking into account the frequency effect. Kilgariff (1995) compiled a lemmatized frequency list of 6,318 words that occur more than 800 times in the 100-million-word British National Corpus (BNC). Words shown in the frequency list are treated similarly to headwords in an English dictionary. This means that a word like *so*, which can be an adverb or a conjunction, has two entries in the list. Kilgariff's frequency list ranks the Additive *so* (conjunctive) in 144th position and *therefore* in 435th, whereas *hence* is in 1925th position. Although words with high frequency "are not necessarily restricted to relaxed speech, it is a fact that informal expressions tend to be high in frequency" (Bolinger, 1972, p. 22). In the case of the three Additive markers, it is likely that *so* and *therefore* will be employed more commonly than *hence* in English conversation. This frequency effect, however, is less straightforward when discussing MDMs, adding to the difficulty that L2 learners face in employing MDMs in accordance with the proper formality context.

To reveal the formality differences among various MDMs, I include frequency data from four corpora, the Cambridge and Nottingham Corpus of Discourse in English (CANCODE), the academic discourse in the British National Corpus (BNC-Acad), the British Academic Spoken English corpus (BASE), and Enronsent2, an e-mail corpus. I then devise the *Weighted Formality Average* method (WFA) to calculate the formality weight of various MDMs. The data suggest that a formality continuum exists among synonymous MDMs. That is, MDMs that suggest the same semantic link can be placed along the continuum, depending on their relative formality weight. For example, language users tend to use *but then again*, *but still*, and *having said that* in

conversation when forming a Concessive link, whereas *despite that* or *in spite of that/this* are preferred in formal writing.

The intertwined relation between the semantics and the context-sensitive feature of MDMs makes MDMs an important area in L2 pragmatic studies. The lack of register awareness in L2 learners' choice of SDMs seems convincing, as different studies have reached the same conclusion. Nevertheless, the lack of data obtained from learner spoken corpora makes this line of argument problematic. The statement that L2 learners are insensitive to register difference will only be verified if researchers interpret the findings from both the written and the spoken learner data. Because learner corpora in the spoken format are not widely available, I seek an alternative by including two experiments: a reaction time task and a multiple-choice discourse completion task. It is hoped that by using both corpus data and experimental designs, a clearer picture of learners' pragmatic awareness in formality or register differences—with regard to the use of MDMs—will emerge.

1.2 Aims of this thesis

Combining corpus methodologies with experiments, the present thesis seeks to offer a full description of how synonymous MDMs are used in various registers, and further, to investigate L2 learners' pragmatic awareness with regard to the use of different MDMs. Specifically, I narrow the scope of pragmatic awareness to the discourse level to see how learners show discourse-level awareness in using MDMs in different registers.

Four stages are involved in this thesis. In the first stage, I examine MDMs in four English corpora that only collect native speakers' (NS) linguistic data. The goal is to provide a full description of MDM distributional patterns in various registers. In the second stage, I examine the same set of MDMs in the Cambridge Learner Corpus (CLC). I describe the patterns of MDMs used by learners with various level of English proficiency in the formal and informal registers. Learner data is further compared with the data from the four NS corpora. The third stage involves two experiments: a reaction time task (RT), and a multiple-choice discourse completion task (MDCT). The variables manipulated in the two experiments are the formality of MDMs, the semantics of MDMs, and English proficiency (NS vs. learners). The goals of the two experiments are to see whether or not learners perceive the formality or stylistic differences in their choices of MDMs in the two tasks, and to seek the possibility of reflecting and quantifying learners' pragmatic awareness. The last stage is to compare the learner corpus data with the experimental results. The purposes are twofold: to see if the experimental results support the corpus findings and to evaluate the experimental approach taken in reflecting L2 learners' pragmatic awareness.

Five research questions will be addressed:

1. Does a formality continuum exist within each semantic relation that distinguishes synonymous MDMs: for example, differentiating *what's more* from *in addition* within the Additive semantic category?
2. Provided that an MDM formality continuum does exist, do learners perceive the formality differences in MDMs and use these markers

accordingly? Or do L2 learners show a similar lack of register awareness in using MDMs, as is the case with SDMs?

3. Does learners' insensitivity to formal register difference apply to spoken/less formal discourse?
4. Does English proficiency affect L2 writers' use of MDMs in terms of quantity, varieties, and types?
5. To what extent can we measure and quantify learners' pragmatic awareness in registers via a reaction-time paradigm and an MDCT device?

1.3 Structure of this thesis

In this chapter, I have described the goals and the research questions of this thesis. I have introduced issues to be addressed in this research: MDMs, Interlanguage Pragmatics, formality and registers, and learners' pragmatic awareness in register difference. In Chapter 2, I will introduce definitions of and concepts related to DMs and interlanguage pragmatics in more detail. Previous research on L2 DM studies and issues that this thesis aims to explore will be reported. Chapter 2 also includes the conceptual frameworks that are to be used in this thesis, and in analysing experimental data.

In Chapter 3, I will provide a detailed description of MDMs through a series of corpus-based studies. Three methods of calculating formality scores, including the one I introduce, the *Weighted Formality Average (WFA)* method, will be compared and evaluated. I will present the formality scores of the targeted MDMs and propose a

formality continuum on the basis of the results obtained from the *WFA* method.

Chapter 4 includes an investigation of learner English in the Cambridge Learner Corpus (CLC). It thus gives a detailed account of how learners use MDMs in formal and informal written discourse. Data obtained from learners at different levels of English proficiency will be compared to see how English ability is related to the quantity and type of MDMs employed.

Chapter 5 presents two empirical studies that employ corpus-informed materials as the testing items. Study 1 explores the possibility of detecting pragmatic awareness of register difference via a reaction-time paradigm; specifically, I manipulate the formality of MDMs, and that of the contexts, to investigate whether a formality mismatch results in a prolonged reaction time. Study 2 aims to detect and to quantify learners' pragmatic awareness via a multiple-choice discourse completion task (MDCT). The two experiments will provide empirical evidence to be compared with the learner corpus data detailed in Chapter 6. Chapter 6 therefore provides a detailed comparison between the two types of data: the learner corpus data and the experimental data. A thorough discussion on learners' pragmatic competence as well as the implications of the findings will be presented. I will also evaluate the feasibility of the experimental approach used in detecting learners' pragmatic awareness. In Chapter 7, I will draw upon findings from both the empirical studies and the corpus-based studies to illustrate the main contributions of this thesis, focusing particularly on learners' pragmatic awareness of register or formality differences. The final sections of Chapter 7 will detail limitations of the approaches taken (corpora and experiments) and identify those areas with potential for future work.

Chapter 2 - Discourse Markers and Interlanguage Pragmatics

Having laid out the thesis structure as well as the main research questions to be addressed, I aim to provide in this chapter a review of studies and theories proposed with regard to DMs within the fields of discourse analysis and corpus-based studies. Findings from corpus-based studies that examine L2 learners' use of DMs will be reported. Literature on pragmatic competence, Interlanguage Pragmatics, and the conceptual frameworks employed will also be reviewed. On the basis of the L2 DM studies, I argue that the neglected area, MDMs, indeed takes a salient role in the study of the development of interlanguage pragmatics.

2.1 What are discourse markers?

In order to make sense of one's discourse, other than employing the right words and grammar, a writer or a speaker has to keep in mind logic and coherence. One of the devices that language users have at their disposal to make their writing or speech coherent is *discourse markers* (hereafter DMs). DMs are words like *but*, *and*, *nevertheless*, prepositional phrases like *on the contrary* and *as a result*, and phrases such as *while I think of it* and *speaking of which*, that are "sequentially dependent elements which bracket units of talk" (Schiffrin, 1987, p 31). Schourup (1999) reviewed various literature on DMs, summarizing DMs as "a more or less open class of syntactically optional, non-truth-conditional connective expressions" (p. 242).

DMs are considered crucial in daily speech and conversation, as they signal the opening or the closing (*right, okay, well and so*), the topic boundary (*oh yeah, so, and*),

and they help in terms of maintaining the on-going discourse (*as I was saying, that is to say, to put it another way*) (Carter & McCarthy, 2006, p. 108). When DMs are examined in written English, however, research usually focuses on the coherence relations that they create. DMs are therefore usually referred to as linking adverbials (Biber et al., 1999) in written discourse. For example, employing DMs like *nevertheless* and *despite* would suggest that a writer is trying to create a Concessive coherence relation.

The most common DMs that Biber et al. (1999) found (their term: linking adverbials) in English conversation are *so, then, anyway* and *though*, whereas it is *however, thus, therefore*, and *for example*, which occur with relatively high frequency in academic prose. Carter and McCarthy (2006) also reported that the most common DMs in everyday informal spoken language are single words (*anyway, cos, fine, good, and great*, etc.), and phrasal and clausal items (*you know, I mean, and mind you*). Markers like *you know* and *right* are also seen as pragmatic devices used as positive politeness strategies to convey solidarity and to establish shared understanding with the addressee (Holmes, 1993).

The brief introduction above has shown that DMs include various components, ranging from single-word markers to multi-word units. Fraser (1990, 1996) argued that the group of DMs consists of heterogeneous constituents such as prepositional phrases (*on top of that, in any case*), adverbs (*nevertheless, however*), and conjunctions (*and, but*). Carter and McCarthy (2006) considered DMs as a lexical category and claimed that to group DMs "in terms of the conventional major word classes (noun, verb, adjective, adverb) is problematic since they stand outside of phrase and clause structures, and they are best considered as a class in their own right" (p. 209). The

heterogeneous constituents of DMs have led to a substantial body of research on DMs in the field of discourse analysis. Various terms are used or coined to refer to the same or similar linguistic phenomena, including *pragmatic connectives* (van Dijk, 1979; Stubbs, 1983), *pragmatic markers* (Fraser, 1988; Fraser, 1990; Schiffrin, 1987), *cue phrases* (Grosz & Sidner, 1986), *discourse connectives* (Warner, 1985; Blakemore, 1987, 1992), *sentence connectors* (Arapoff, 1968), *text connectives* (Vande Kopple, 1985), *logical connectives* (Crismore et al., 1993; Hyland, 1998) and *conjunctive elements* (Halliday & Hasan, 1976), *pragmatic particles* (Chalker & Weiner, 1994), and *pragmatic force modifiers* (Nikula, 1996), to name just a few.

Despite the various terms, there are some characteristics of DMs that clearly distinguish them from other linguistic constituents. Schourup (1999) put forward seven features that are mostly commonly attributed to DMs—connectivity, optionality, orality, multi-categoriality, non-truth-conditionality, weak clause association, and initiality. In the following section I will briefly summarize the first four features, which are directly related to this thesis.

2.2 Features of DMs

2.2.1 Connectivity

The widely accepted characteristic of DMs is that they signal the coherence relationship of on-going and foregoing discourse units. The use of different markers results in a change of the connectivity expressed. Consider (1) below:

(1) It has been raining for days.

- a. **On the contrary*, we decide to go swimming.
- b. *Nevertheless*, we decide to go swimming.
- c. *As a result*, we decide to go swimming.
- d. ?We decide to go swimming.

Even though the proposition of (1b) to (1d) remains the same, i.e., raining and going swimming, the logical level elicited by the three markers varies (Schiffrin, 1987). The use of *nevertheless* in (1b) creates a Concessive tie, whereas the use of *as a result* in (1c) forms a Causal tie. *On the contrary* in (1a) is an odd case, as readers or listeners cannot find a cohesive tie in the discourse. Without a proper DM like (1d), although it is less comprehensible, (1d) is still considered acceptable if both speakers and interlocutors share the contextual assumption to some extent (as is suggested by 1c).

2.2.2 Optionality

As the word optionality suggests, the occurrence of DMs in discourse is optional. With the first feature, connectivity, in mind, this suggests that the omission of a DM does not alter the semantic relationship suggested. This is shown by (2):

(2) It has been raining for days. We decide not to go swimming.

In the case of (2), the cohesive tie, a Causal relation, is established despite the fact that a DM is no longer seen. That is, without the use of DMs, speakers and writers still maintain the level of communication with pragmatic knowledge or via other

linguistic devices, such as the use of deictic expressions and anaphoric pronouns (Redeker, 1991), or ellipsis, repetition, and substitution (Halliday & Hasan, 1976).

Another meaning that optionality denotes is that the use of DMs does not affect the syntactic structure of its host sentence (Fraser, 1988). It should be noted that even though the use of DMs is optional both semantically and syntactically, DMs “are commonly said to reinforce, or clue the interpretation intended by the speaker” (Schourup, 1999, p. 232). This is particularly evident when testing L2 learners’ comprehension. Chaudron and Richards (1985) found that the use of DMs in lectures helps L2 learners’ comprehension in general (see Section 2.5 for details).

2.2.3 Orality

Treating orality as one feature of DMs seems to suggest that most of the DMs tend to occur mainly in speech. Schourup admits this seemingly biased view by stating that the reason for orality as one feature of DMs is simply that “early work on DMs focused predominately on conversational items like *well* and *oh*” (1999, p. 234). This feature does not rule out the fact that some DMs (*moreover, consequently*) tend to occur mainly in written discourse.

Instead of seeing orality as one general feature, I propose that it is the *formality* that explains the fact that some DMs tend to be found largely in speech whereas others occur mostly in written discourse. As the corpus data reveals (see Chapter 3), a few DMs occur equally frequently in speech and writing.

2.2.4 Multi-categoriality

As explained earlier, what constitutes DMs does not come from one syntactic class. Rather, DMs include adverbs (*however, anyway*), prepositional phrases (*on top of that, in addition*), clauses (*I mean, you know*), interjections (*oh, gosh*), and verbs (*look, say*), etc. Multi-categoriality refers to this specific nature. Note that the categories that Schuourup defined refer to the syntactic categories only. For the purpose of the study, I extend the notion of multi-categoriality to semantic category. DMs have another distinctive feature that is also termed multi-functionality (Hatch, 1992; McCarthy, 1991). When the same DM is used to signal a different semantic relationship, this DM shows its multi-functional feature. For example, the MDM *on the other hand* can be a Concessive, a Listing, or a Contrastive marker (Quirk et al., 1985), indicating three semantic categories depending on the context.

The above four characteristics are directly related to the theme of this thesis. In terms of connectivity, this thesis aims to explore the extent to which L2 learners use MDMs to relate one discourse unit to another. In spite of the optionality feature and all the possible linguistic devices to use instead of DMs, one question worthy of considering is: do L2 learners recognise this optionality, or is it simply due to the limited linguistic repertoire of L2 learners that this optionality becomes less optional? With regard to orality, I intend to examine whether and to what extent L2 learners recognise the formality of different MDMs, and therefore use MDMs appropriately in various registers. The last feature of the multi-categoriality (multi-functionality) of DMs is crucial, as examining MDMs in both NS and L2 data requires that the step of disambiguation be taken in order to distinguish polysemous MDMs such as *on the other hand* and *at the same time*.

Another often mentioned characteristic of DMs is their syntactic positions in sentences or utterances. In this thesis, I intend to examine the distributional patterns of MDMs in various registers. That is, of particular interest is how the formality of context affects the choice of MDMs. The syntactic position becomes less relevant. The position information is therefore neither included in the discussion nor in the data analysis stage.

Schuurup (1999) noted that a great deal of emphasis in DM studies, including the aforementioned research, has been placed on single-word discourse markers (SDMs) in spoken discourse (*oh, well, y'know, and I mean* in Schiffrin, 1987; *however, still, anyway, and incidentally* in Lenk, 1998; *so* in Blakemore, 1988). Comparatively fewer studies examine the roles DMs play in written discourse (Vande Kopple, 1985; Hyland, 1998; Sanders et al., 1992). In the later studies that examine DMs in written discourse, the emphasis remains centred around SDMs like *nevertheless* and *however*. Although some DM studies include a few MDMs, MDMs still receive only marginal attention. One exception is Siepmann's (2005) contrastive corpus-based study on MDMs, in which they are referred to as *second-level discourse markers*.

2.3 Two main accounts of DMs

Theoretical studies that explore the roles of DMs are carried out based on two main frameworks: the relevance and the coherence approach. Although the two accounts have assigned different roles to DMs, the distinction or the dispute between the two is not within the scope of this thesis. Since this thesis is developed based on

the coherence approach, a brief paragraph below introduces the relevance framework, followed by a detailed description of the coherence account.

The relevance account (Blakemore, 1987, 1992, 2002; Blass, 1990; Iten, 1998; Wilson & Sperber, 1993; Sperber & Wilson, 1986) considers discourse connectives (DMs in this thesis) as indicators to show relevance in discourse. Advocates of this framework give DMs a crucial role in interpreting utterances because the relevance of utterances is constrained by the use of various DMs, facilitating listeners'/readers' arrival at the intended cognitive effect. For example, the use of *but* in (3) facilitates the inferential process, creating the assumption that John will not travel in the summer.

(3) Mary is going away this summer but John has to re-sit his statistics class.

In the coherence framework (Halliday & Hasan, 1976; Schiffrin, 1987; Fraser, 1988, 1990, 1999; Mann & Thompson, 1986; Sanders et al., 1992), the most salient function of DMs in a discourse is to signal the coherence relations between the foregoing and the following discourse. Halliday and Hasan (1987) proposed that what makes a text semantically well-formed is whether or not a cohesion is discerned in it. DMs are one of the five coherence devices they discussed in creating coherence in texts. When DMs relate two adjacent discourse units and create coherence, such coherence is referred to by Schiffrin (1987) as *local coherence*. After examining various DMs, conjunctions and phrases, Fraser suggested a broader sense of coherence that DMs create—"a DM imposes on S2 [segment 2] a certain range of interpretations, given the interpretation(s) of S1 and the meaning of the DM, a topic to be discussed

subsequently" (1999, p. 942). How DMs create both the local and the broader sense of coherences will be examined in the present study.

Under this coherence framework, DMs are therefore categorized into different coherence relations, such as the Additive coherence relation (*furthermore, in addition*), and the Contrastive coherence relation (*but, on the contrary*), as (3) illustrates. Nevertheless, theorists have not reached a consensus as to how many coherence relations there are and therefore they propose different sets of coherence relations, with two coherence relations suggested by Grosz and Sinder (1986), four coherence relations by Halliday and Hasan (*ibid.*), six general semantic categories proposed by Biber et al. (1999), seven reported by Quirk et al. (1985), sixteen suggested by Mann & Thompson (1986), and over four hundred coherence relations reported by Hovy and Maier (1995). Just as some DMs exhibit multi-functionality, as mentioned in section 2.2.4 (Hatch, 1992; McCarthy, 1991), some DMs occur under more than one semantic category. For instance, Mann and Thompson (1986) point out that *but* marks either Contrastive (their term: thesis-antithesis) or Concessive coherence relations.

On the basis of the coherence relations suggested by Halliday and Hasan (1976), Quirk et al. (1985), Biber et al. (1999), and Carter and McCarthy (2006), five semantic categories of MDMs were chosen for this study: Additive, Appositive, Causal, Concessive, and Contrastive (see Section 3.5). It should be noted that although Additive and Enumerative are deemed by Quirk et al. and Biber et al. as one broad category, the Enumerative category is fundamentally used to show the order of pieces of information, whereas the use of Additive markers explicitly points out the similarity of two items or objects. In this study, Additive MDMs do not include Enumerative

markers, such as *first of all*, and *to begin with*. For a similar consideration, Concessive is treated as a different semantic category from Contrastive, even though some researchers consider both under the Contrastive semantic category (Quirk et. al., 1985; Biber et al., 1999).

2.4 DMs in context—a pragmatic perspective

Other researchers examine DMs from a pragmatic perspective. Rieber (1997) used *but* to discuss the pragmatic meaning it brings to texts or utterances. Following this, Blakemore (2002) explored further the uses of *but*, *however*, and *yet*, suggesting that DMs carry pragmatic meaning because “they do not contribute to the truth conditional content of the utterances that contain them” (p. 12). In a similar vein, Fraser (1990) proposed that DMs are “extremely useful guides for clarifying a speaker’s communicative intention” (p. 390). Expanding the pragmatic notion by bringing in registers and styles, Stubbe and Holmes (1995) conducted a study of social and stylistic variation in the use of DMs and found that markers like *eh*, *I mean* and *you know* are linked to working and middle class speech styles in New Zealand English. Sinclair and Coulthard (1975) investigated classroom discourse and identified that the frequent markers (their term: frame) used by teachers in this register are *right*, *well*, *good*, and *O.K.*

It is evident that the registers or the styles of discourse have an influence on the choice of DMs language users make. Let us use the following three MDMs, *despite that*, *having said that*, and *at the same time* to illustrate this point further. Although all the three MDMs suggest a Concessive relation and are synonymous to some extent, what

differs among them is not a straightforward answer. A more intriguing question is: what is the deciding factor for a language user to prefer one synonymous MDM to another when the semantic nuance is not easily perceived? Consider (4a) to (4e).

(4) John's folks paid for the damage for the car accident;

(a) *having said that*, they were pissed off.

(b) *at the same time*, they were pissed off.

(c) *still*, they were pissed off.

(d) *?nevertheless*, they were pissed off.

(e) *?despite that*, they were pissed off.

Most competent English speakers would agree that (4a) to (4c) sound appropriate, whereas they might find odd the use of *nevertheless* in (4d) and *despite that* in (4e). An obvious trait we can detect from the context is the vocabulary used in (4). *Folks* and *pissed* are colloquial words that people use in informal settings. *Nevertheless* and *despite that* are more formal and tend to be used in formal, written discourse. That is, the DMs used in (4d) and (4e) do not seem to match the informality shown in the context. What differentiates (4c) from (4d), which are both SDMs, and from (4a-b) and (4e) which are MDMs, is the level of formality, which is embedded within the notion of pragmatic consideration.

The odd feeling that arose from hearing my former classmate use *hence* in a casual conversation illustrates the inappropriateness that one feels when the markers employed do not correspond to the formality of the context or the register.

Considering the former classmate's use of *hence* semantically, he achieved his goal; yet, he seemed not to pay attention to the pragmatic consideration and was oblivious to the register differences between a formal setting (academic writing) and an informal situation (casual conversation). Such a mismatch between form and function, i.e., using the formal form *hence* in an informal setting to fulfil communicative function, is observed in the field of Second Language Acquisition.

However, how do L2 learners perceive the difference between various levels of formality among MDMs? Unlike overt grammatical features, which can be pinpointed via morphology or syntax in most languages, formality features on the phrasal level that suggest either a formal or an informal register tend to be less obvious. One exception is observed in those formality features that occur in speech acts. Making a request with bi-clausal forms (*would it be possible to VP?* or *would you mind if I VP?*), presumably, carries more formal weight than a mono-clausal form request (*Can you VP?*). Takahashi (2001) observed that Japanese learners of English tended to use mono-clausal forms in making requests, and employed fewer bi-clausal forms. Nevertheless, if we look at formality features on a lexical level, one cannot easily perceive the formality difference, and this is certainly the case with MDMs.

Even though the extent to which the choice of a particular DM indicates the formality of registers and styles remains intuitively obvious, such register-oriented usage difference does not seem to be discussed extensively in the literature, with the exception of the following corpus-based studies. Carter and McCarthy (2006) discussed the differences between spoken and academic English and listed DMs that tend to be used in the two main areas. Biber et al. (1999) examined the corpus data

and proposed that DMs (their term: linking adverbials), which indicate result/inference, are commonly seen in conversation and academic prose. Altenberg (1986) used the term 'difference coefficient' (*diff. coeff.*) to propose a formality divergence between spoken and written Contrastive SDMs and MDMs, which together he called *links*. Despite Altenberg's successful attempt to create a formality scale for Contrastive markers, his data was retrieved from two relatively small samples of corpora (100,000 words each). Although Carter and McCarthy (2006) as well as Biber et al. (1999) included MDMs in their investigation, the focus of these works is still on SDMs. Only MDMs that occur frequently in academic prose, such as *in addition* and *to sum up*, were mentioned in the two research works. What is needed is a more detailed exploration that not only brings in large-scale corpora but also examines MDMs in the less formal register. This study explores this lacuna by using various spoken and written corpora, and seeks to offer a fine-grained description of MDMs with regard to their level of formality when used in different contexts.

2.5 Corpus-based studies of L2 learners' use of DMs

With the rise of various types of corpora, such as Standard English corpora (like the British National Corpus) and learner corpora (like the International Corpus of Learner English, and the Cambridge Learner Corpus), researchers have made use of the accessible data in evaluating and describing learner language, including collocations (Granger, 1998; Liu, 2002; Nesselhauf, 2005), phrasal verbs (Waibel, 2007), and formulaic language (Myles et al., 1998; Fitzpatrick, 2005). DMs are also widely examined via a corpus-based approach.

Hyland (1998) collected twenty-eight research articles from refereed journals in four academic disciplines and examined the metadiscourse used, which included hedges, sequencing items, and DMs (his term: logical connectives). He argued that metadiscourse were means used by writers to construct appropriate contexts, to perform illocutionary acts, and to achieve perlocutionary effects. Hyland (2004) later examined metadiscourse in a corpus of doctoral and masters' dissertations by Hong Kong students and concluded that DMs are crucial for advanced students in their academic writing in "uncovering something of the rhetorical and social distinctiveness of disciplinary communities" (2004, p. 133). The finding is consistent with others who found metadiscourse to be a key index of good ESL and native speaker student writing (Intaraprawat & Steffensen, 1995; Cheng & Steffensen, 1996). That is, employing a variety of appropriate DMs not only increases the intelligibility of a discourse, but they also "serve a rhetorical purpose" (McCarthy, 1991; McCarthy & Carter, 1994, p. 50). This is not to say that DMs used by L2 learners are flawless. Two problems that Hinkel (2002, 2004) observed with regard to how L2 writers use DMs are, in comparison with NS writers, that L2 writers use too many markers (her term: sentence transitions), and that the use of DMs "does not necessarily make the L2 academic writing cohesive or the information flow easy to follow" (2004, p. 292).

Milton and Tsang (1993) conducted a corpus-based study of Hong Kong students' use of DMs (their term: connectors) and discovered that students tend to overuse logical markers. To use the DMs in a more native-like fashion, they suggested that students "must possess, beyond semantic and syntactic knowledge of the language, a sensitivity to register as well" (1993, p. 216). Field and Yip (1992) found that the

sentence-initial position is commonly adopted by L2 writers, whereas L1 writers use the markers more in the non-initial position. Similarly, Granger and Tyson (1996) examined 108 DMs (their term: connectors) in the writing of French learners of English in the International Corpus of Learner English (ICLE). Other than overuse of Additive and Appositive DMs, they found misuse of DMs like *in fact* and *indeed*, which is due to the L1 interference, *en fait*¹. They also reported learners' tendency to place DMs in sentence-initial position. Learners' overuse of DMs is reported by other researchers who investigate L2 learners with various L1 backgrounds, including Chen's (2006) study on Taiwanese learners of English; Narita, Sato and Sugiura's (2004) study on Japanese learners of English; Field and Yip's (1992) research on Cantonese speakers of English; Bikeliene's (2008) study on Lithuanian Learners of English.

Another often reported conclusion from studies of DMs used by L2 learners is that L2 learners generally lack pragmatic awareness in employing proper DMs in accordance with the context (Field & Yip, 1992; Granger & Tyson, 1996; Altenberg & Tapper, 1998; Lorenz, 1999). Specifically, L2 learners seem to use more informal DMs in their essay writing. Altenberg and Tapper (1998) raised two problems that learners face—using DMs (their term: connectors) without discrimination, and the lack of sensitivity in using DMs in various registers and discourse types. Other studies that report learners' tendency to use spoken features in written texts include Aijmer (2002), Neff *et al.* (2007), Narita and Sugiura (2006), and Gilquin and Paquot (2008).

Tankó looked at DMs (his term: adverbial connectors) in argumentative essays written

¹ According to Granger and Tyson, *en fait* is used to mean 'Here I am in the text' and is often used as a stylistic enhancer. The English connector *in fact*, however, does not function like this.

by Hungarian learners of English and concluded that “the writers’ register awareness appears to be appropriate for the production of formal discourse” (2004, p. 178). Such argument is contradictory to other L2 DM studies reported above in which L2 learners were found to be insensitive to the register differences. Variables that might contribute to the contrasting conclusions include the learner corpus size and the different English levels possessed by the participants of these studies. Most of the literature reported has suggested that L2 learners face a challenge in using DMs in accordance with the academic writing style. Even though Tankó’s finding contradicts the aforementioned studies, both Tankó and other researchers have focused exclusively on L2 academic writing. What seems to be neglected is how L2 learners employ DMs in a less formal register.

Although studies have reported that experienced L2 writers use more DMs (Barton, 1993; Intaraprawat & Steffensen, 1995; Hyland, 2004), such findings again only show the link with regard to the use of DMs in writing. Only limited investigation has been made in exploring roles of DMs in learner speech (Fung & Carter, 2007). Fung and Carter conducted a corpus-based investigation of DMs used in a pedagogical setting by NS and by Hong Kong learners of English. Based primarily on the data collected, they proposed a categorical framework for DM classification and argued for the vital role of DMs in L2 spoken English, particularly in facilitating understanding and communication, and thus providing “learners with a sense of security in L2 [English]” (p. 433).

Another study that investigates the correlation of DM use to L2 acquisition is the research done by Chaudron and Richards (1985). They conducted an experiment to

see how the use of DMs in lectures affects ESL learners' comprehension. They prepared four versions of a recorded lecture, a Baseline (no DMs) version, a Micro version (with DMs like *then, at that time, after this*), a Macro version (using phrases like *what I am going to talk about today*) and a Micro-Macro version (a combination of the Micro and the Macro versions). The results of comprehension tests showed that the Micro-Macro version assisted the participants' comprehension the most, whereas the use of DMs only (the Micro version) did not assist participants in keeping track of the lecture content. Although their study tested to what extent L2 learners' comprehension of a lecture is affected by not only the use of DMs but also other phrases, the fact that the design included DMs suggests the important role DMs assume in L2 learning.

Despite a growing number of L2 DM studies, the unit of analysis is still SDMs. Although such a tendency echoes the practice of mainstream discourse studies, it also signals the attention needed to other types of DMs (see Schuourup's claim in Section 2.2.4). Issues related to L2 learners' use of DMs are raised mainly based on the SDM data. If we look at the Concessive coherence relation again, there are indeed other legitimate markers within this category, which includes *despite this/that, but at the same time*, and *but then again*. Questions such as 'how do L2 learners use MDMs in their output?' and 'do L2 learners also exhibit a lack of sensitivity to registers?' are not addressed. Further, the extent to which the exploration of this aspect could reveal learners' interlanguage development remains under-explored.

The above statement is not to suggest that MDMs are not studied. Rather, MDMs have received ample research interest in the study of formulaic language and are often

referred to as prefabricated chunks (Wray, 2000, 2002). Studies in this regard explore how L2 learners use formulaic sequences in their language output (Schmitt, 2004). Examining MDMs from this angle focuses either on whether or not L2 learners retrieve these multi-units as a whole, or on how the acquisition of multi-word units progress over time. The coherence relationships created via the use of MDMs are not the focus of study. In other words, the study of MDMs under the term 'formulaic language' has missed the essential nature that MDMs possess: the cohesive ties. This thesis examines how the cohesive links of MDMs, when combined with various levels of formality, are reflected in L2 language output.

Another frequently discussed issue in L2 DM studies is the breakdown of coherence relations caused by the inappropriate use of DMs (Wikborg & Björk, 1989; Mauranen, 1993; Altenberg & Tapper 1998; Hinkel, 2002, 2004). Literature reviewed thus far suggests that reasons behind the cases of misuse include meaning confusion (Granger & Tyson, 1996), formality mismatch (Field & Yip, 1992; Granger & Tyson, 1996; Altenberg & Tapper, 1998; Lorenz, 1999), or L1 interference (Granger & Tyson, 1996; Hinkel, 1999; Scollon & Scollon, 2001). None of the above studies, however, pursue further the pragmatic awareness that lies behind the misuse case. In addition, researchers have not looked at what types of DMs tend to be used more erroneously than others. In other words, would some coherence relations tend to be more complex to express in terms of the uses of DMs? Sanders et al. (1992) designed a labelling experiment in which participants first learned the differences between, for example, Additive and Contrastive relationships, and were given pairs of sentences at a later stage. Participants were asked to read the paired sentences and to label what semantic

relationship each pair suggests. The result showed that of all the semantic categories Sanders et al. examined, Concessive coherence relation seemed to be the least identifiable and least agreed among all the participants. The level of labelling difficulties within various relations could in turn suggest the order of acquisition of the coherence relations; however, they did not extend this notion to L2 learners. I will pursue the matter further in the learner corpus (Chapter 4) and the two experiments in Chapter 5.

2.6 Pragmatic competence

Although the emphasis of the reviewed L2 DM studies is not on L2 learners' pragmatic awareness, the ability to perceive the formality of context and to choose DMs accordingly falls into the category of pragmatic competence. In what follows, I will discuss the broader concept of communicative competence from which pragmatic competence derives, and present the conceptual frameworks adopted in the later stage of data analysis.

Other than the traditional linguistic repertoire, which includes the knowledge of grammar and vocabulary, the notion of *pragmatics* in SLA has also received substantial attention. Pragmatics explores "the factors that govern our choice of language in social interaction and the effects of our choice on others" (Crystal, 1987, p. 120) and it is also "the study of those relations between language and context that are grammaticalised or encoded in the structure of language" (Levinson, 1983, p. 9). In a recent work, O'Keeffe, Clancy, and Adolphs (2011) reviewed various studies on pragmatics and proposed that pragmatics is "concerned . . . with accounting for the processes that give

rise to a particular interpretation of an utterance that is used in a particular context” (p. 2). Another definition that O’Keeffe et al. (2011) quote is from Christie (2000), who defined pragmatics as a theoretical framework that “can account for the relationship between the cultural setting, the language user, the linguistic choices the user makes, and the factors that underlie those choices” (2000, p. 29). For L2 learners, however, developing this ability in pragmatics, and understanding contextual linguistic choices, can be a “daunting learning task” (Kasper & Roever, 2005, p. 317).

Putting together the definitions above gives rise to the concept of *appropriateness* (Hymes, 1972; Campbell & Wales, 1970; Dewaele, 2008). Van Dijk (1977) also put forward the significance of context in his discussion of pragmatics and proposed *appropriateness* in evaluating pragmatic success. That is, employing the right words/phrases in accordance with the context has become another central issue in language acquisition. Such an idea, construed within the notion of *communicative competence* (Hymes, 1972), has inspired researchers in exploring learner language, particularly in situated uses. In order to arrive at the understanding of pragmatic competence, which is conventionally considered to be a component of communicative competence, it is useful to consider in some detail some of the literature that probes communicative competence, namely that of Hymes (1972), Canale and Swain (1980), Canale (1983), Bachman (1990) and Celce-Murcia *et al.* (1995). Some theoretical constructs consider communicative competence from the pedagogical point of view, and others from the language testing perspective; yet of interest here is the notion of pragmatic ability discussed in each model. I will therefore give prominence to the pragmatic component in each model.

2.6.1 Hymes's communicative competence

Hymes's (1972) communicative competence originated from his objection toward Chomsky's (1965) distinction between competence and performance, in which the word *competence* exclusively refers to grammatical knowledge. Hymes discussed the idea of communicative competence by incorporating sociocultural factors, which he presumed should be borne in mind by taking into account the following four types of knowledge when *acceptability* is at issue: whether (and to what degree) something is formally *possible*; whether (and to what degree) something is *feasible* in virtue of the means of implementation available; whether (and to what degree) something is *appropriate* (adequate, happy, successful) in relation to a context in which it is used and evaluated, and finally, whether (and to what degree) something is in fact done, actually *performed*, and what this doing entails (p.281, his emphasis). In other words, Hymes conceived the idea of communicative competence as a grammatical, psycholinguistic, sociocultural, and probabilistic system (Canale & Swain, 1980, p. 16). The knowledge of whether (and to what degree) something is *appropriate* reflects pragmatic consideration in the sociocultural context. In addition to the grammatical knowledge one possesses, Hymes believed that to say someone knows a language, this person must know "when to speak, when not . . . , what to talk about with whom, when, where, in what manner" (1972, p. 277). Thus, signing off a business-related email with 'Cheers' might sound friendly yet unprofessional when the email is sent to a business contact, and starting a conversation at a supervision meeting with a casual 'What's up?' rather than a regular greeting such as 'How are you?' would sound awkward.

2.6.2 Canale and Swain's communicative competence

Canale and Swain (1980) developed a theoretical framework of communicative competence, which they proposed to be adopted in second language teaching. In their framework, communicative competence comprises three competences: grammatical competence, sociolinguistic competence, and strategic competence. Grammatical competence, as the term suggests, includes knowledge of lexical items, rules of syntax, phonology, and morphology, etc. Sociolinguistic competence consists of two rules: sociocultural rules of use, and rules of discourse. The concept of *appropriateness* mentioned earlier or the idea of pragmatic competence is included in the category of sociocultural rules in that "certain propositions and communicative functions are appropriate within a given sociocultural context", and "appropriate attitude and register or style are conveyed by a particular grammatical form within a given sociocultural context" (1980, p. 30). The third competence—strategic competence—is made up of verbal and nonverbal communication strategies that language users need when communication breaks down. Canale (1983) added a fourth component into the model: discourse competence. Discourse competence, according to Canale, is language users' ability to decode linguistic forms and rules in various kinds of discourse. Communicative competence, as a result, is defined as "the underlying systems of knowledge and skill required for communication" (Canale, 1983, p. 5). The model proposed by Canale and Swain (1980) and Canale (1983) is illustrated in Figure 2.1.

Although the notion of pragmatic competence was included in Canale and Swain's (1980) model, it was not until Bachman's framework that pragmatic competence started to take a prominent role in the discussion of communicative competence.

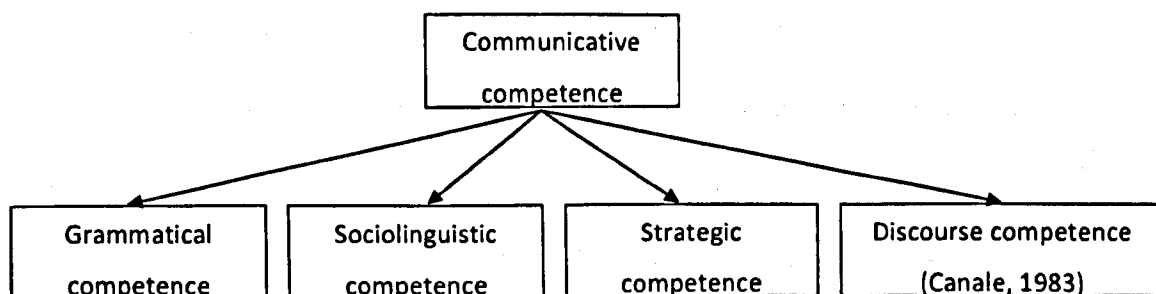


Figure 2.1 Communicative competence constructed in Canale and Swain (1980) and Canale (1983)

2.6.3 Bachman’s communicative competence

In addressing communicative language ability in relation to language-testing, Bachman (1990) proposed a thorough framework of communicative competence, which is represented in a tree structure (see Figure 2.2). The tree structure might give readers an impression of hierarchy, but Bachman stressed the dynamic feature of language use, and stated that the components interact with each other. It is the “very interaction between the various competencies and the language use context that characterizes communicative language use” (Bachman, 1990, p. 86). Communicative competence, in his view, comprises two components: *organizational* and *pragmatic* competence. Organizational competence further includes grammatical and textual competence, which parallels Canale’s (1983) discourse competence. Pragmatic competence, on the other hand, has two dimensions: illocutionary competence and sociolinguistic competence. Illocutionary competence is related to various speech acts (Searle, 1969), such as refusals, requests, and apologies, performed in signalling

speakers' intents. Sociolinguistic competence, alternatively, refers to the knowledge of how to use language functions appropriately in a given context. Sociolinguistic competence is "the sensitivity to, or control of the conventions of language use that are determined by the features of the specific language use context" (1990, p. 94). Bachman considered four abilities in defining sociolinguistic competence: sensitivity to differences in dialect or variety, sensitivity to differences in register, sensitivity to naturalness or the native-like way (Pawley & Syder, 1983), and ability to interpret cultural references and figures of speech.

Bachman's subdivisions of pragmatic competence resemble Leech's (1983) classification of pragmatics. Leech subdivided pragmatics into sociopragmatics and pragmalinguistics, with the former focusing on socially appropriate language use and the latter on the linguistic strategies and items needed to achieve certain speech intentions.

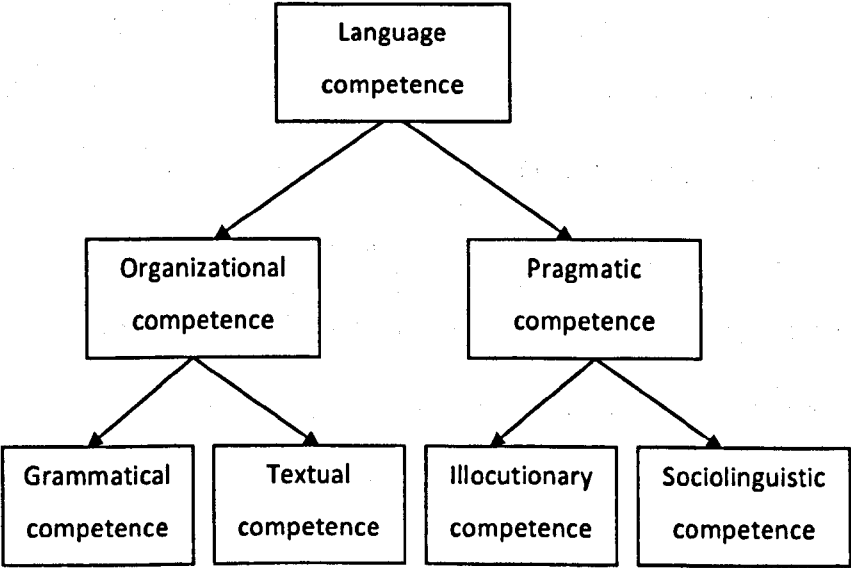


Figure 2.2 Bachman's framework of communicative competence

Crucial for our discussion of pragmatic competence in Bachman's model is the salience he gave to language users' *sensitivity to differences in register*, which is the major issue that this thesis intends to explore. Bachman adopted Halliday, McIntoch, and Strevens' (1964) three aspects in addressing differences in register, namely, *field of discourse*, *mode of discourse*, and *style of discourse* (1964, p. 90). Considering the three aspects together, a competent language user will be able to perceive the register differences, and to speak, to write, or to respond in an appropriate way. Following this line of thinking, the purpose of examining learner corpus data (Chapter 4) as well as conducting the two experiments (Chapter 5) is to explore whether or not L2 learners consider the differences between the formal and informal registers and write or respond in an appropriate way.

2.6.4 Celce-Murcia, Dörnyei, and Thurrell's communicative competence

Another well-cited framework of communicative competence is Celce-Murcia *et al.*'s (1995) model, which is pedagogically-motivated, extending the constructs of Canale and Swain (1980), and those of Bachman (1990). Having perceived the problem of separating discourse competence from sociolinguistic competence (see Schachter, 1990 for a detailed critique), and the hierarchical typology of the previous two models, Celce-Murcia *et al.* proposed their model of communicative competence via the concept of a pyramid in which the five constituents/competences—linguistic, strategic, sociocultural, actional, and discourse—are interrelated (see Figure 2.3).

From Figure 2.3, it is clear that discourse competence is placed in the central position where other components come together to shape the discourse. The cycle

surrounding the pyramid represents strategic competence, which compensates for any deficiencies that any of the other competences might have, or allows speakers to communicate strategically. Of particular interest here are actional competence and sociocultural competence. Actional competence is “conceptualized as competence in conveying and understanding communicative intent by performing and interpreting speech acts and speech act sets” (Celce-Murcia et al., 1995, p. 9), similar to Bachman’s illocutionary competence. When L2 learners are in discussion, Celce-Murcia *et al.* claim that this actional competence is equal to Kasper & Blum-Kulka’s (1993) interlanguage pragmatic competence. Sociocultural competence, on the other hand, is defined by how speakers take into account social and cultural contexts, presenting their messages appropriately. This is the extension of Bachman’s sociolinguistic competence.

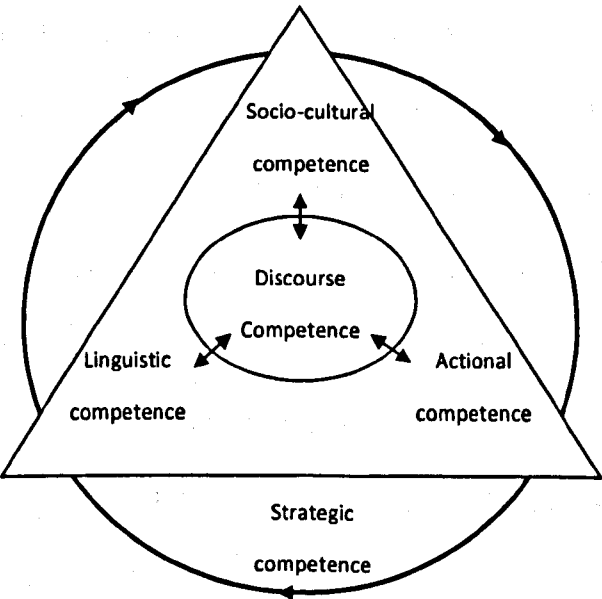


Figure 2.3 Celce-Murcia et al.’s schematic representation of communicative competence

Four types of factor are included in Celce-Murcia et al.'s sociocultural competence: social contextual factors, stylistic appropriateness factors, cultural factors, and non-verbal communicative factors. The second group, stylistic appropriateness factors, is similar to Bachman's sensitivity to differences in register. From the pedagogical perspective, Celce-Murcia et al. made clear that L2 learners have to be made aware of stylistic variation. That is, learners have to take notice of degrees of formality and differences in field-specific registers (1995, p. 23).

Note that when reviewing literature on actional competence (or illocutionary competence), and sociocultural competence (or sociolinguistic competence) in L2 acquisition, one can easily see an overwhelming number of studies on how various speech acts are performed via illocutionary competence, such as compliments in Rose & Ng (2001), and Lorenzo-Dus (2001), requests in Hassall (1997), Li (2000), Rose (2000), and Ellis (1992), refusals in Felix-Brasdefer (2004), and apologies in Trosborg (1995), Schauer (2006), *inter alia*. One exception that brings in the element of sociolinguistic competence in speech acts studies is Garcia's (2004) study, in which she examined how L2 learners' pragmatic awareness is perceived via various speech acts. Unlike speech acts that have been well represented in cross-cultural and interlanguage pragmatics research (see Section 2.7), research on the sociolinguistic aspect, particularly on how L2 learners perceive the style or register differences, seems to be relatively limited.

From the four models introduced, we perceive a progressive refinement of what constitutes and defines pragmatic competence. Discussing linguistic competence by itself is inadequate to fully describe communicative competence. It is also clear that

constructs of communicative competence are increasingly giving more consideration to pragmatic competence. Bialystok (1993) adopted a cognitive approach in her definition of pragmatic competence, which I use to summarize this discussion of what defines pragmatic competence. She suggested that pragmatic competence is the interpretation of meaning in contexts, and the symbolic representation of this performance is a mapping between form and social context, rather than between form and meaning. The development of pragmatic competence parallels the development of the "resource of equivalents from which selections can occur" and "the richer the repertoire, the greater would be the pragmatic competence" (1993, p. 51)

Describing pragmatic competence within the field of Second Language Acquisition gives rise to interlanguage pragmatics. In the following section, I will narrow down the scope of pragmatics to that of interlanguage pragmatics and present the theoretical frameworks that motivate this study.

2.7 Interlanguage pragmatics

Interlanguage pragmatics (ILP), as the term suggests, originates from two disciplines, with Second Language Acquisition (SLA) on the one hand, and pragmatics on the other (Kasper & Blum-Kulka, 1993). As a result, ILP is defined as "the study of nonnative speakers' use and acquisition of L2 pragmatics knowledge" (Kasper, 1996, p. 145). Barron (2002) defined L2 pragmatic competence as the "knowledge for realizing particular illocutions, knowledge of the sequential aspects of speech acts, and finally, knowledge of the appropriate contextual use of the particular language's linguistic resources" (p. 10). Alcón & Jordà (2008) explored pragmatic awareness in L2

acquisition and defined pragmatic awareness as “the conscious, reflective, explicit knowledge about pragmatics” (p. 193). They further argued that the knowledge “of those rules and conventions underlying appropriate language use in particular communicative situations and on the part of members of specific speech communities” (p.193) should be included in developing L2 learners’ communicative competence.

2.7.1 Methodologies in ILP

The majority of ILP studies adopt data elicitation devices, such as discourse completion tasks/tests (DCT), questionnaires, multiple-choice questions, and role play. These elicitation designs are usually presented to participants with carefully-planned wording, with the goal of reducing any variation that might undermine the designs. Observational, authentic data collection is therefore comparatively rare (Bardovi-Harlig & Hartford, 2005). Kasper and Dahl (1991) reviewed the methods of data collection employed in thirty-nine ILP studies and reported that only two of these studies used observational data. To meet this need, the two experiments (see Chapter 5) designed to test L2 learners’ pragmatic awareness include corpus-informed materials.

Another commonly observed characteristic in ILP is that an overwhelming body of research focuses on various speech acts (as reported in Section 2.6.4), regardless of other facets of pragmatic competence discussed in the previous section.

Bardovi-Harlig (2010) gave a thorough review of studies of ILP and stressed the need for research in this under-explored area. In her attempt to investigate how the methodologies of L2 pragmatics and the related research designs have evolved in the

past three decades, Bardovi-Harlig (2010) examined 152 journal articles, reporting in detail that the nature of replicating research methodologies has resulted in “the dominance of certain speech acts, elicitation tasks, and population” (p. 243). In terms of study areas, Bardovi-Harlig proposed that “researchers should build on existing research to study the unstudied” (2010, p. 243). L2 learners’ awareness of or sensitivity to the formality differences in different registers and contexts belongs to the category of the *unstudied*, and this is the very topic this thesis aims to examine.

Much of the research in the ILP literature follows the methodology commonly used in the social sciences and includes either cross-sectional or longitudinal designs to examine interlanguage development in various stages. Cross-sectional designs collect data from two or more groups of a sample on the basis of, for example, L2 proficiency differences (Olshtain & Blum-Kulka, 1985; Rose, 2000; Taguchi, 2002, 2005; Lee, 2010). Longitudinal studies, on the other hand, focus on the same group of participants over a certain period of time so that any systematic development in pragmatic performance is observed (Achiba, 2003; Bardovi-Harlig & Hartford, 1993).

2.7.2 Factors in developing interlanguage pragmatics

Despite the different methodologies adopted, the literature suggests some decisive factors in interlanguage development. Kasper and Roever (2005) reported that *input* (length of stay in the target language environment), *noticing and understanding* (whether or not attention is paid to relevant input features), *L2 proficiency*, *L1 pragmatic transfer*, and *individual differences* are the main variables that influence the acquisition and development of interlanguage pragmatics. Although

some of the factors might be more influential than others in cases such as L2 proficiency in Rose (2000), Taguchi (1999), and Bardovi-Harlig & Dornyei (1998), and length of stay in Bouton (1992), and Achiba (2003), together these factors decide to what extent L2 learners produce appropriate language in accordance with the contexts.

2.8 The adopted theoretical frameworks

As ILP is a hybrid discipline of SLA and pragmatics, theoretical frameworks employed in ILP research tend to be used as a *post hoc* explanation of the results rather than motivating the studies (Kasper & Rose, 2002). Kasper and Rose (2002) reviewed various studies on pragmatics development in L2 learners and discussed five theoretical perspectives, which derive from two origins: cognitive and socially-oriented perspectives. As this thesis deals with pragmatic awareness, I will focus exclusively on the cognitive-oriented perspective and introduce the theoretical frameworks that are adopted in this thesis: the noticing hypothesis and accommodation theory.

2.8.1 Noticing hypothesis

Schmidt (1990, 1993) proposed the *noticing hypothesis* in his discussion of the role of consciousness in learning a second language. The noticing hypothesis claims that, in order for language acquisition to take place, linguistic features have to be noticed at the outset. I borrow this idea of noticing a linguistic feature, and extend the noticing from the linguistic feature to the level of formality inherent in the context. The concept

of noticing will therefore be applied to the analysis of the experimental data (Chapter 5). Schmidt distinguished noticing from understanding and defines *noticing* as “registering the simple occurrence of some event” while *understanding* refers to the “recognition of a general principle, rule, or pattern” (1993, p. 26). *Noticing* decides “what linguistic material is stored in memory”, and *understanding*, on the other hand, is about “how that material is organised into a linguistic system” (1993, p. 26). That is, while *noticing* refers to “surface level phenomena and item learning”, understanding is about the “deeper level(s) of abstraction related to meaning, system learning” (Schmidt, 1995, p. 29). Employing the noticing hypothesis in this thesis, I use *noticing* to refer to learners’ attention to various MDMs in written or spoken discourse, and *understanding* as learners’ register awareness when they need to choose markers in accordance with the formality of the discourse. In other words, while it is likely that in a reading task with the linking markers missing, an L2 learner would choose, for example, *in addition* for a formal Additive passage, it is quite likely that the same learner would choose the more colloquial marker *what’s more*, depending on whether or not it is the *noticing* or the *understanding* that leads to the decision. This parallels Schmidt’s claim that in order to develop pragmatic competence, “one must attend to both the linguistic forms of utterances and the relevant social and contextual features with which they are associated” (2001, p. 30). The noticing hypothesis is therefore adopted in examining L2 learners’ pragmatic competence.

2.8.2 Accommodation theory

The basic concepts of accommodation theory (Giles, 1973; Giles et al., 1973; Street & Giles, 1982; Giles & Coupland, 1991) are that “a speaker could be expected to converge her or his speech towards that of a listener in certain respects” or alternatively, diverge to “symbolize the desire to reduce intimacy” (Coupland, 2007, p. 62). The theory is used to “account for diverse contextual processes that impinge on the selection of sociolinguistic codes, styles and strategies and their interactional consequences” (Coupland, 2010, p. 21). Giles (1973) argued that Labov’s (1966) claim of language variation, in which the alternative forms of speech do not occur at random but are decided by certain factors (social identity, historical position and stylistic level, etc.), can be “reinterpreted, at least in part, as having been mediated by interpersonal accommodation processes” (Giles et al., 1991, p. 5). As it provides a framework in which the variation in communication performance or the linguistic choices can be analysed from the cognitive point of view, accommodation theory has been applied to various disciplines (Meyerhoff, 1998) and to different groups of interlocutors, with various social and institutional contexts (courtroom communication, Gnisci 2005; police-civilian interactions, Giles et al., 2007). The majority of research that adopts accommodation theory is mainly focused on conversational or face-to-face communication. A few studies employ the theory in other models, such as in email communication (Bunz & Campbell, 2004), and in telephone conversation (Buzzanell et al., 1996).

Accommodation theory is used in this thesis to describe readers’ accommodative behaviour in the two studies (reported in Chapter 5). When reading academic, formal passages, competent readers presumably would accommodate to the complex syntax

and low frequency words of the texts provided in the reading process. When asked to find a missing marker for such passage, they are more likely to choose a marker that is associated with formal writing style. When reading informal discourse segments, on the other hand, readers will then expect simplified grammar and lexical items that tend to occur in a conversation. Their choice for the missing marker, in this case, should be an informal one. Both cases are seen as examples of a convergent process. In the accommodation theory, where interlocutors choose not to converge, the implication is that they wish to maintain the distance between themselves and the people they are conversing with. When readers in the two experiments do not accommodate to the formal context and choose an informal MDM, for example, this is not considered as a divergent process; this will be seen as an example of insensitivity to the level of formality. In other words, noticing in the formality as well as in the semantic category of the missing marker has to take place before the accommodative process occurs.

2.9 Summary and conclusion

I have reviewed the literature on L2 DM studies and reported issues that have not yet been explored, such as the pragmatic awareness that lies behind learners' misuse of DMs, and also the marginalized unit of analysis, MDMs. In addition, I have reported conceptual frameworks for understanding pragmatic competence and introduced the two theories adopted in analysing experimental data: the noticing hypothesis and accommodation theory. Large corpus-based studies mainly focus on identifying the frequently occurring SDMs. Although Biber et al. (1999) maintained that DMs (their

term: linking adverbials) realized in prepositional phrases (MDMs) occur relatively frequently in academic prose, they only introduce the most common ones. The subtle semantic nuance among synonymous MDMs is not pursued.

Before I explore the learner data, it is important to offer a detailed description of MDMs in terms of their distributional patterns in various registers. This will be done by examining MDMs in four reference corpora in the next chapter. Specifically, I will focus on pinpointing the influence that contexts have upon the choice of MDMs. A formality continuum will be introduced on the basis of the formality weights obtained from three calculation methods.

Chapter 3 - Towards a Description of Multi-word Discourse Markers in Context

Having laid out the problems that L2 learners face in using DMs as well as a review of the literature on DMs, which exclusively focuses on SDMs, in this chapter I present the unit of study in this thesis: multi-word discourse markers (MDMs). The goal of this chapter is to explore the distributional pattern of MDMs in the following four corpora—the academic section of the British National Corpus (BNC-Acad), the British Academic Spoken English corpus (BASE), the Enron E-mail corpus, and the Cambridge and Nottingham Corpus of Discourse English (CANCODE). Taking into account the frequency information from the four corpora, I devise a Weighted Formality Average method (*WFA*) for calculating the formality weight of the targeted MDMs. I compare the *WFA* method results with two other existing formality calculation methods—Altenberg’s difference coefficient (*diff. coeff.*) method and Brooke, Wang, & Hirst’s formality score (*FS*) method. Other than classifying MDMs into various semantic categories, I propose a more refined typology of MDMs in terms of formality, a formality continuum, which is created on the basis of the formality weights calculated for each targeted MDM. This formality difference separates, for example, *despite that* from *but then again* which are conventionally clustered to be Concessive markers. The goal of this chapter is to describe the distributional pattern of MDMs based upon the formality consideration. Once such a pattern is established, it will have its pedagogical contribution when this pattern is compared with the MDMs used by L2 learners.

3.1 MDMs and registers: situated language use of MDMs

In Chapter 2, I have shown that, like grammar and vocabulary, the use of MDMs varies depending on the formality suggested by the context. Formality can be used to refer to a situation where there is an “increased structuring and predictability of discourse” (Irvine, 1979, p. 774). In a similar vein, Richards, Schmidt, and Kendricks defined ‘formal speech’ as “a careful, impersonal and often public mode of speaking used in certain situations and which may influence pronunciation, choice of words and sentence structure” (2002, p. 209). However, as Heylighen & Dewaele (1999) pointed out, this definition only conceptually suggests what a situation is like, and such a criterion is still non-linguistic. The linguistic dimension of formality has been described as the language used to reflect either the “degree of intimacy” or the “degree of seriousness” (Rubin 1968).

Levels of formality shown in any situated language use are related to *genre*, *register*, or *styles*. Reid (1956) first used the term ‘register’, and the concept of register was later further illustrated in the work of Ferguson (1977), Ellis and Ure (1969), Andersen (1990), and Biber and Finegan (1994). Register is used to refer to the variety of language used for a situational purpose and setting (Carter & McCarthy, 2006; Richards et al., 2002). As the clarification between registers, styles, and genre is not the focus of this thesis (see Lee, 2001 for a detailed discussion), I will simply use the word ‘register’ throughout the thesis when I refer to style, or genre.

Various facets of linguistic difference exist among registers, and some criteria are proposed for the purpose of identifying various registers (Biber, 1994; Duranti, 1985). Trudgill used linguistic varieties in his description and proposed that “registers are

usually characterized solely by vocabulary differences; either by the use of particular words, or by the use of words in a particular sense" (1983, p. 101). In his discussion of registers, Finegan looked at language use in relation to formal and informal social occasions and used the four words *pickled*, *high*, *drunk*, and *intoxicated* to illustrate the "situational continuum along which forms of expression may be arranged," with formality and informality as opposite poles of this continuum (2008, p. 319). Biber and Conrad (2001) reviewed various literature and summarized characteristics of registers which include the participants, their relations, their attitudes toward communication, the setting, the level of formality, and the channel of communication. Following these characteristics, for example, a professor may employ precise language and use prescriptive grammar when addressing the Dean of a university (e.g. *How do you do?*); the same professor might instead greets his friends and family with a more friendly, informal phrase, *How's it going?* In other words, register variation is reflected via the use of vocabulary, syntax, and morphology. Important research that links language users' selection of DMs within various registers has been pursued by Brizuela, Anderson, and Stallings (1999). They demonstrate how the uses of DMs vary when their participants talk with lower-status roles (children) or with higher-status roles (parents, teachers) and conclude that DMs can be seen as an indicator of registers.

Defining register and genre—or the formality of context—involves various factors. For the purpose of this thesis, I adopt the "culturally recognizable category" of genres (Adolphs, 2008, p. 80) when discussing contexts and formality. In Adolphs' discussion of genre and text-type, she considered genres as folk-categories or

culturally recognised activities. Following this, a conventionally accepted genre for a formal context would be academic discourse, whereas a culturally recognised genre for informal context would be daily, casual conversation among friends. Leech et al. (1982, p. 140) suggested that the difference between written and spoken language, however, should not be seen as 'watertight subcategories of mode' but instead there exists a certain level of overlap between them. They suggest that conversation in a pub is typical speech, whereas a book about grammar would represent typical writing. Radio conversation, a television advertisement, a lecture and a sermon would take a neutral position. The intertwined relationship between situated language use and the formality of the context and language channels can be illustrated in Figure 3.1.

Labels in Figure 3.1 like *Written English* and *Spoken English*, *Formal setting* and *Informal setting* might look like two contrastive points, yet the division is in fact a continuum that is gradable. This continuum is represented as a line with an arrowhead on each end. The different exemplary contexts given in Figure 3.1 are prototypical types of genres. It does not imply, for example, that emails are absolutely typical of informal written genres, nor that a supervision meeting must be very formal. Berkenkotter and Huckin (1993) discussed this changing nature of discourse, describing genres as "dynamic, constantly (if gradually) changing over time in response to the sociocognitive needs of individual users" (p. 481).

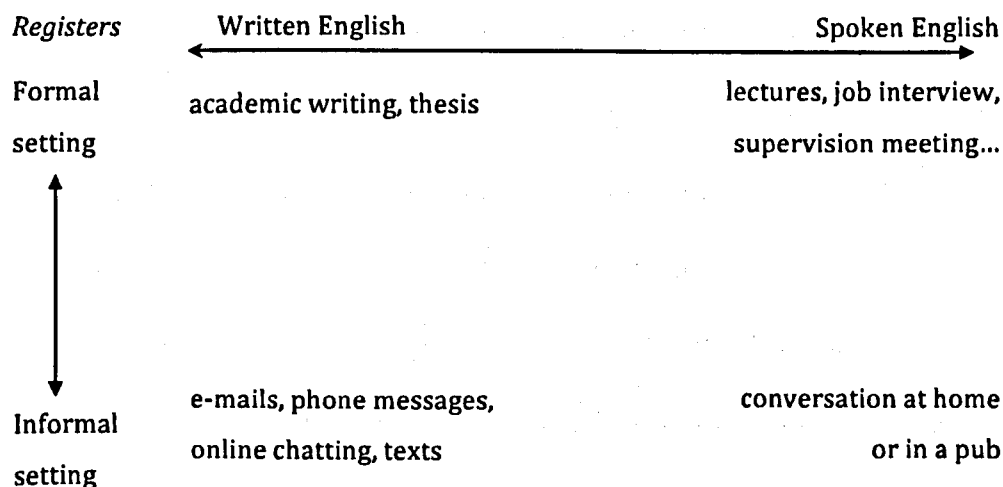


Figure 3.1 The interaction of language use and the registers

With the figure in mind, I propose that various MDMs occur in accordance with their situated uses in expressing coherence relations. The formality continuum of MDMs, I assume, is ascertainable by examining different language in use. Before I describe the creation of the formality continuum, I will introduce a relevant, yet early attempt to create such a continuum, Altenberg's 1986 research.

3.2 Altenberg's formality scale

In his attempt to find the usage differences of Contrastive DMs (which he called *links*) in spoken and written English, Altenberg (1986) carried out a corpus-based study and proposed a formality divergence between spoken and written DMs. He examined the distribution of the Contrastive markers in the London-Lund Corpus of Spoken English (LLC), and in the Lancaster-Oslo/Bergen Corpus of British English (LOB), proposing a formality scale. Spoken and written are the two extremes on the

scale, with neutral items in the middle. To quantify the differences perceived via frequency counts, he adopted Hofland & Johansson's (1982) difference coefficient (*diff. coeff.*) method. *Diff. coeff.* is calculated by considering the raw frequencies of the target items in both spoken and written corpora. In his study, each DM (or link) is assigned a *diff. coeff.* which ranges from +1 to -1. A marker with a positive value shows its salience in the spoken sample. A negative score, on the other hand, suggests a marker's representation in the written corpus. The formula is shown below (Altenberg, 1986, p. 14):

$$\frac{\text{Freq: Spoken} - \text{Freq: Written}}{\text{Freq: Spoken} + \text{Freq: Written}}$$

Altenberg considered *diff. coeff.* as a stylistic measure and proposed three categories on the basis of the scores obtained: *style-restricted* DMs, which are confined to either speech or writing, *style-biased* markers, which are less restricted but “at least twice as common in either variety,” and *style-neutral* markers, which show little or no preference in either direction (Altenberg, 1986, p. 17). In Table 3.1, DMs in (1) and (5) belong to the *style-restricted* group, DMs in (2) and (4) are *style-biased* markers, whereas DMs in (3) are *style-neutral*.

Table 3.1

Stylistic differentiation of Contrastive DMs proposed in Altenberg's study (1986, p. 18)

Links (DMs)	DIFF. COEFF.	BIAS
anyway	+1.00	(1) Specific to speech
though (conjunct)	+0.86	
(but) then	+0.83	
after all	+0.36	(2) Promoted in speech
but	+0.31	
whereas	+0.27	
even if	0.00	(3) Style-neutral
in any case	0.00	
still	-0.09	
nevertheless	-0.14	
instead	-0.20	
on the other hand	-0.29	(4) Promoted in writing
though (subordinator)	-0.32	
at any rate	-0.38	
although	-0.42	
yet	-0.49	
while	-0.79	(5) Specific to writing
however	-0.90	

Despite Altenberg's successful attempt to create a formality scale for Contrastive DMs, the data shown in Table 3.1 is retrieved from two small samples of corpora, with 100,000 words each. Further, Altenberg only examined one type of DM—words or phrases that show contrasts. His Contrastive DMs, however, include both Concessive DMs and the Antithetic conjuncts. Some frequently used Concessive MDMs like *having said that* and *this/that said* are absent from his data. What is needed is a more detailed exploration that not only brings in large-scale corpora but also examines other types of DM. Nevertheless, I will include Altenberg's *diff. coeff.* in this chapter, along with other statistical analyses to perceive the register/style difference between synonymous MDMs.

3.3 The four corpora

Due to the fact that the current corpora are compiled either as spoken or written corpora (rather than as formal or informal databases), an academic written English corpus will be the main source for formal written English, whereas a corpus that archives spoken, casual conversation will provide data needed for the informal spoken source. In total, four corpora are used to adduce formality: the Cambridge and Nottingham Corpus of Discourse in English corpus (CANCODE), the academic discourse in British National Corpus (BNC-Acad), the British Academic Spoken English corpus² (BASE), and the Enronsent2 e-mail corpus. Table 3.2 shows the nature and the size of these corpora, followed by a brief introduction to each.

² The corpus was developed at the Universities of Warwick and Reading under the directorship of Hilary Nesi and Paul Thompson. Corpus development was assisted by funding from BALEAP, EURALEX, the British Academy and the Arts and Humanities Research Council.

Table 3.2

Summary of the four corpora included in this chapter

Corpus	Register	Words
BNC-Acad	Written Eng. (academic)	2,074,185
BASE	Spoken Eng. (academic)	1,186,290
Enronsent2	Written Eng. (informal)	9,826,382
CANCODE	Spoken Eng. (conversational)	4,859,392

3.3.1 The Cambridge and Nottingham Corpus of Discourse in English (CANCODE)

The CANCODE corpus (Carter & McCarthy, 1995) is a nearly five-million-word database that has collected spontaneous conversation or dialogues taking place at different locations (such as shops, workplaces or homes) with different functions (chatting, asking for information, discussion) in the United Kingdom. All the spoken data was recorded initially, and then transcribed by professional analysts.

3.3.2 The academic discourse in British National Corpus (BNC-Acad)

The British National Corpus (BNC), a 100 million-word collection of samples of British English (both written and spoken language), is considered a balanced database as it collects language samples from various sources. The academic section of the BNC (BNC-Acad) consists of academic prose from periodicals and non-fiction books from a range of different subject areas.

3.3.3 The British Academic Spoken English (BASE) Corpus

The BASE corpus that is used in this study is the version accessible on Sketch Engine (Kilgarriff et al. 2004), which consists of 160 lectures recorded in two universities in the UK. As in the BAWE corpus, the lectures are mainly from four disciplines: Arts and Humanities, Social Sciences, Life Sciences, and Physical Sciences. Although it is a spoken corpus, the register tends to be academic. Because it is not as colloquial as conversation taking place between friends, the BASE corpus is considered as a spoken, formal data source.

3.3.4 The Enronsent2 corpus

The Enron Corpus dataset, which is available to download online (Klimt & Yang, 2004), is a large collection (0.5M messages) of e-mail messages released during the Federal Energy Regulatory Commission (FERC) investigation into the Enron Company in the United States. The email messages archived are mainly communication between colleagues and to customers. Although the e-mail messages do not fully represent an informal writing style, the language style in e-mails chosen tends to be informal or speech-like (Baron, 2003). Various researchers have made use of this e-mail dataset in email research, and in natural language processing (Dredze et al., 2008; Lampert et al., 2008; McCallum et al., 2007; Peterson et al., 2011). I include a subset of the Enron e-mail dataset as representative of a less formal register on the formality continuum. The subset database is referred to as the Enron Sent Corpus, tailored specifically for research in Corpus Linguistics (Styler, 2011). Although Styler had removed message headers, quoted messages, and HTML messages, I discovered that the Enron Sent

Corpus still includes masses of duplicated messages, resulting from the forwarding nature of e-mail communication. As the purpose of this chapter is to pinpoint the formality differences, which requires accurate frequency counts, I took an additional pruning step to eliminate those duplicated messages³. The version that I used in this thesis is referred to as the Enron Sent 2 Corpus (Enronsent2).

Linking the four corpora to Figure 3.1, the four corpora represent the data source for the four intersections. The BNC-Acad corpus provides formal, written data, such as academic writing, theses, and journal articles. The BASE corpus offers formal, spoken language samples, such as job interviews and supervision meetings. The Enronsent 2 corpus provides informal, written data sources, whereas the CANCODE corpus provides informal, spoken language samples, such as conversation taking place in a pub.

To search the BNC-Acad corpus, I used the *creating a subcorpus* function provided by the search interface of Sketch Engine (Kilgariff et al. 2004). I obtained access to the British National Corpus via the Cambridge International Corpus⁴. Under the 'creating a subcorpus' option, the BNC corpus is sub-divided into four categories: academic, fiction, spoken and written. It is through the sub-classification that the search of BNC-Acad was made possible. To search the data archived in the BASE corpus, I used the open corpora search interface provided by Sketch Engine. After downloading the Enron e-mail corpus, I used Notepad++ software in the search for the MDMs. A line break occasionally occurred during the Notepad++ search and therefore additional

³ For this pruning step, I am indebted to Dr Tsao Nai-Lung, who used a python script to delete the duplicates in the EnronSent Corpus.

⁴ I am an approved individual researcher under the English Profile research project (www.englishprofile.org/) and am granted access to the Cambridge International Corpus

manual examination was undertaken. For the search of the CANCODE corpus data, I used the corpus investigation interface developed by Cambridge University Press.

3.4 Operational definition of multi-word discourse markers (MDMs)

All the MDMs included in this thesis are defined as linguistic units that stand on their own syntactically, and are separated with a comma, or commas, when occurring in the sentence/utterance initial or median position. Although their existence is optional, once used they clearly signal a coherence relation that links a previous discourse to a following discourse. Although the positions of the MDMs are not of concern in this thesis, the majority of the target MDMs tends to occur in the sentence-initial or median position. Multi-word units like *I mean* and *y'know*, which do not show an explicit coherence relation, are therefore excluded from this thesis. In addition, the term 'DMs' is considered a neutral term, and therefore, when used in this thesis, refers to both MDMs and SDMs.

Some MDMs, like *despite this/that* and *in spite of this/that*, might occur "with the pronoun (*that*) filled in with a modified form of S1 [the previous segment], and occurs in a reverse pattern" (Fraser, 1998, p. 303). This is exemplified in the following example (1a) through (1c).

- (1) a. Carrie left her office late. *Despite this*, she arrived at the dinner on time.
b. *Despite leaving her office late*, Carrie arrived at the dinner on time.
c. Carrie arrived at the dinner on time, *despite leaving her office late*.

In these examples, only the use of *despite this* in (1a) functions as a DM and this is the standard that will be followed in data collection.

3.5 The target MDMs

On the basis of four reference books—Halliday and Hasan (1976), Biber et al. (1999), Quirk et al. (1985), and Carter and McCarthy (2006)—five semantic categories (Additive, Appositive, Causal, Contrastive, Concessive) and twenty-five types of MDM were chosen (see Table 3.3). Note that DMs can co-occur with coordinating conjunctions (Carter & McCarthy, 2006, p. 284; Quirk et al., 1985, p. 640), and therefore cases like *at the same time* and *but at the same time* are treated as two tokens of one type. Likewise, *despite that* and *despite this* are two tokens of the same type: *despite that*. When an SDM co-occurs very frequently with *and* or *but* (for example, *but still*), such a combination is seen as a multi-word unit and is considered as an MDM. Although the four reference books include neither *and also* nor *but still*, these two are legitimate, frequently used markers. They are therefore included in this thesis.

There are *marker-like constructions* that are not included in the four reference books. A good example is the apposition-like phrases such as *what I'm saying is* and *what this means is*, which typically occur at the beginning of an utterance/sentence. Quirk et al. (1985) subcategorized four semantic levels under Appositive markers. Appositive MDMs like *that is to say* and *in other words* are clustered under the label Equivalence, and signal appellation, identification, designation, and reformulation (1985, p. 1308). The apposition-like phrases, *what I'm saying is* and *what this means is*,

can also be used to signal the above relations. These two phrases are therefore included in the Appositive MDMs in Table 3.3. I follow these four relations as the criteria when I examine all the occurrences of *what I'm saying is* and *what this means is* in the four corpora. That is, the Contrastive use of *what I'm saying is*, which normally occurs after *no* or *but*, is not included.

Table 3.3

*Target MDMs shown in the semantic categories (The symbol * refers to MDMs that are not from the four reference books)*

Semantic category	Multi-word discourse markers (MDMs)
Additive	what's more, not only that, in addition, on top of that, besides that, add to this, *and also
Appositive	that is to say, in other words, *what I'm saying is, *what this means is
Causal	as a result, because of that, as a consequence, for that reason
Contrastive	on the other hand, on the contrary, in contrast
Concessive	despite that, having said that, but then again, in spite of that, at the same time, on the other hand, *but still

In total, twenty-five types of MDMs were searched in the four corpora (see Appendix 1 for a full list of all tokens of the twenty-five MDMs).

3.6 Methods of data analysis

I adopt both qualitative and quantitative approaches in analysing the retrieved data.

3.6.1 Qualitative approach

The four corpora examined in this chapter are not semantically tagged. Some MDMs are multi-functional, and therefore polysemous. An example given by Bell (2004) is that *on the other hand* can be a Contrastive marker, a Concessive tie, and a Listing device. That is, a key word in context (KWIC) search of such an MDM retrieves concordance lines that include all the occurrences, regardless of the sense difference. In addition, the less frozen structure such as *because of that/this* retrieves both the connective use of *because of that* and the instances where *because of* is used in prepositional clauses (*because of that problem*). All the concordance lines were examined manually to weed out false positive data.

3.6.2 Quantitative approach

3.6.2.1 Frequency cut-off point and the screening process

For the quantitative approach, I started with the most fundamental information in corpus-based studies: raw frequency. To claim that the observed raw frequencies of a word/phrase in two different corpora are significantly different from one another, a frequency cut-off point is usually set. Such cut-off point is based on a normalized frequency (McEnery et al., 2006): the frequency count per million words (pmw). This is to provide a common base so that the comparison is justified when made among

corpora that vary in size, as is the case in the four corpora reported in this chapter. Different frequency cut-off points have been reported—twenty times pmw by Hyland (2008), forty times pmw by Biber et al. (2004), and by Biber (2006), ten times pmw by Biber et al. (1999), and four times pmw by Carter and McCarthy (2004). It should be noted, however, that the main purpose of the above studies is to look for emerging patterns of lexical bundles (Biber et al., 1999) or multi-word clusters. When judging whether or not a four-word or five-word chunk like *you know what I*, or *know what I mean* is salient, the cut-off frequency becomes crucial. This is not the case with frozen chunks like MDMs, as their linguistic salience has been identified. The meaning of raw frequency in this chapter, therefore, is to serve as a practical anchor in perceiving the register differences. As a result, I adopt a very liberal frequency cut-off point in collecting the data. That is, a very low occurrence of an MDM (for instance, with a frequency of one) will still be included at the first stage, followed by a screening process.

The screening process is used as a safeguard over any pitfalls that might be caused by the liberal frequency cut-off point. Altenberg took the same precaution in his 1986 study, in which he disregarded DMs that are “too infrequent to be stylistically revealing” (1986, p. 15). When the total frequency numbers from the two samples are less than ten, Altenberg excluded them in the calculation of *diff. coeff.* This *too infrequent* phenomenon is also observed in the four corpora. As Altenberg’s two samples (100,000 words each) are comparatively smaller than the four corpora in this chapter, I set a stricter screening process. When the normalized frequencies of one

MDM from the four corpora are less than five, I exclude such data from the final analysis.

3.6.2.2 Additive smoothing—the *add one* rule used for zero frequency

Zero occurrence of an expression might be explained via the nature of the corpus investigated, the sparseness of a word or a phrase, or the size of the corpus. At times, the zero occurrence phenomenon might be expected, and this predictability provides insights into the issue to be addressed; in other cases, the underdispersion might lead to problems, particularly in statistical analysis (Gries, 2011: online). Much literature has been dedicated to this aspect (Gale, 1995; Gries, online; Gries, 2008). In investigating the formality differences of MDMs in spoken and written corpora, it is intuitively likely that very informal MDMs have very low frequency or even zero-occurrence in a formal, written corpus. That is, such zero-occurrence is expected and meaningful. The meaningfulness, however, would disappear if the number zero were brought into the four statistical measures described below, as the zero frequency implies the probability of zero. To rectify this, I adjusted the zero frequency by adopting one smoothing technique, the *add one* rule (Manning & Schütze, 1999; Kilgariff, 2009), which is commonly employed by corpus linguists. Despite the fact that *add one* has been criticized for its failure in dealing with unseen items in the training corpus or in the reference corpus (Gale & Church, 1994), the purpose of this chapter is not to uncover the sparseness of any n-grams, and therefore, *add one* is used to restore the meaning represented by the number zero.

To apply the *add one* rule, I add the value one to all the frequencies of MDMs that are found to have zero occurrences in any of the four corpora. There are two such cases when retrieving the raw frequency counts: the Additive MDM *add to this*, and the Appositive MDM *what I'm saying is*. I will use *what I'm saying is* to illustrate how this is done. The Additive MDM *what I'm saying is* occurs in the BASE corpus, the Enronsent2 corpus, and the CANCODE corpus, but it does not appear in the BNC-Acad corpus. As mentioned earlier, the zero-occurrence of the colloquial marker-like phrase *what I'm saying is* in BNC-Acad is expected, as BNC-Acad archives very formal, academic discourse. Considering zero in the statistical analysis, multiplying or dividing the number zero will result only in the elimination of the meaning that the zero carries. Applying the *add one* rule, the frequency counts for the four corpora become 1 for the BNC-Acad corpus, 11 for the BASE corpus, 4 for the Enronsent2 corpus, and 6 for the CANCODE corpus. These frequencies are then normalized on the basis of per million words, followed by the four statistical measures.

3.6.2.3 Methods in calculating formality scores

The formality difference of MDMs in spoken and written registers can be quantified via three methods of calculation: difference coefficient method (*diff. coeff.*), simple formality score (*FS*) (Brooke et al., 2010), and the method that I introduce in the following section, the Weighted Formality Average (*WFA*). In what follows, I briefly describe how each method is employed in calculating the formality weight.

DIFFERENCE COEFFICIENT (*diff. coeff.*)

I have briefly introduced Altenberg's (1986) study, in which he examined Contrastive discourse markers by looking at the frequency distribution in spoken and written English. Altenberg used raw frequency in the formula (see Section 3.2), as his items were from two samples of texts, each of which amounted to 100,000 words. As the corpora included in this chapter vary in size, I will use normalized frequencies in the formula. The *diff. coeff.* method is applied to cross-corpus comparisons (Table 3.4); yet only the *diff. coeff.* arrived at from the comparison of the BNC-Acad and CANCODE corpus is compared with the other two formality measures. The *diff. coeff.* scores obtained from other cross-corpus comparisons will be applied to the Weighted Formality Average Method (WFA), described below.

BROOKE, WANG, AND HIRST'S SIMPLE FORMALITY SCORE (FS)

In Brooke et al.'s (2010) attempt to automatically determine the formality of individual lexical items, they proposed various methods including both simple formality measures and hybrid methods. Of interest and relevance to this thesis is one of their simple formality measures that is similar to, but more sophisticated than, the *diff. coeff.* method. By taking the frequency (n) of a word from a relatively formal corpus, the same word's frequency (m) from an informal corpus, and the ratio of the informal corpus to the formal corpus (N), Brook et al. proposed the following formula in calculating the formality score (FS):

$$FS = -1 + 2 \frac{n}{m \times N + n}$$

The *FS* score arrives at ranges from 1 to -1. Unlike the meaning of 1 (informal index) and -1 (formal indication) in the calculation of the *diff. coeff.* score, a positive *FS* of a word in this method suggests that the word is relatively formal, while a negative *FS*, on the other hand, means that it is a colloquial word. To make valid the comparisons of the formality measures, all the *FS* values yielded from the above formula will go through a process of transformation. All the positive *FS* will be converted to negative and negative *FS* to positive.

Note that Brooke et al.'s *FS* measure was devised to calculate the formality weight for single words, rather than multi-word units like MDMs; however, as the literature on linguistic formality either takes a macro-level perspective, which focuses on the broad issues such as genre and style (Heylighen & Dewaele, 2002; Herring & Paolillo, 2006), or embarks on a micro-level analysis, which places emphasis on individual lexical items (Brooke et al., 2010), very little is known in terms of measuring the formality of multi-word units. As a result, Brooke et al.'s simple *FS* measure is still included and the method will be compared with the other two measures.

WEIGHTED FORMALITY AVERAGE METHOD (*WFA*)

The Weighted Formality Average method (*WFA*) that I devise is a slight departure from Altenberg's (1986) *diff. coeff.* method. The formality difference that exists between the formal and the informal register, as explained earlier, is perceived as a continuum. The two extreme ends of the linguistic sources—the very informal spoken dataset and the very formal written corpus—constitute the main scaffolding in creating the continuum. I argue, however, that the comparisons made between the two

contrastive corpora do not constitute a sufficient condition. The formality continuum will only provide a more accurate picture if the distributional patterns of the targeted MDMs in other corpora are included in finding the formality weights. In other words, the *WFA* method takes into account the frequency counts from the four corpora.

Two *dimensional differences* are considered in the calculation of weighted formality: formality (formal, informal) and language mode (spoken, written). Take the comparison made between the CANCODE corpus and the BNC-Acad corpus (pair 1 in Table 3.4), for example. Because the former corpus includes informal spoken data and the latter represents formal written discourse, the two corpora differ entirely both in formality and in mode. In this case, the dimension difference factor is 2. On the basis of the dimension difference factor 2, a heuristic estimate of the influence of this pair's *diff. coeff.* score is 2 times stronger than the *diff. coeff.* scores obtained from other pairs. When the comparison is made between the Enronsent2 corpus and the CANCODE corpus, the dimension difference factor becomes 1.5. The two different language modes contribute to the number 1. Although I categorize the Enronsent2 corpus as an informal, written dataset, the corpus includes both formal business e-mails between companies and informal, personal communication among friends. It is because of this complex nature that the formality difference between Enronsent2 and CANCODE is defined as 0.5. That is, the *diff. coeff.* arrived at from this comparison is only 1.5 times stronger in terms of its impact on the formality calculation. The same rationale applies to the BASE corpus. Even though the BASE corpus consists of academic spoken English, unlike the BNC-Acad corpus, in which the formal notion is intuitively straightforward, when the BASE corpus is compared with the BNC-Acad

corpus, the dimension difference factor is 1.5, with the language mode contributing a value of 1, while the less absolute formality contributes to the remaining 0.5. In other words, the dimension difference factor varies depending upon the corpora being compared. Table 3.4 shows a summary of the dimension difference factors calculated on the basis of the described mechanism.

Table 3.4

The cross-corpus comparisons and the dimension difference factors

Corpora compared	diff. coeff. code	Language mode	Language formality	Dimension difference factor
Pair 1. BNC-Acad	d.c.1	written	formal	2
CANCODE		spoken	Informal	
Pair 2. Enronsent2	d.c.2	written	informal/formal	1.5
CANCODE		spoken	informal	
Pair 3. Enronsent2	d.c.3	written	informal/formal	1.5
BASE		spoken	formal	
Pair 4. BNC-Acad	d.c.4	written	formal	1.5
BASE		spoken	formal (informal?)	
Pair 5. BASE	d.c.5	spoken	formal (informal?)	0.5
CANCODE		spoken	Informal	
Pair 6. BNC-Acad	d.c.6	written	formal	0.5
Enronsent2		written	informal/formal	

The dimension difference factors will be included in the calculation of the *weighted formality average score (WFA)*, which is arrived at via the following formula:

$$WFA = \frac{(d.c.1 \times 2 + d.c.2 \times 1.5 + d.c.3 \times 1.5 + d.c.4 \times 1.5 + d.c.5 \times 0.5 + d.c.6 \times 0.5)}{(\Sigma \text{DimensionDifferenceFactor})}$$

Let us take *what's more* as an example. The normalized frequencies (NF) of this MDM are 1.45 in the BNC-Acad corpus, 2.47 in the CANCODE corpus, 3.37 in the BASE corpus, and 1.12 in the Enronsent2 corpus. The calculated *diff. coeff.* scores for the six possible comparisons are shown in Table 3.5.

Table 3.5
Results of cross-corpus comparisons of what's more

Corpora compared	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6
<i>diff. coeff.</i>	0.26	0.38	0.50	0.40	-0.15	-0.13

Putting the *diff. coeff.* scores in the *WFA* formula, the formality score for *what's more* is 0.31.

$$WFA = \frac{(0.26 \times 2 + 0.38 \times 1.5 + 0.5 \times 1.5 + 0.4 \times 1.5 + (-0.15 \times 0.5) + (-0.13 \times 0.5))}{7.5}$$

= 0.31

As explained earlier, a positive value suggests the marker's preferred occurrence in the spoken register. A value like 0.31 shows that *what's more* tends to occur relatively frequently in spoken English.

3.7 Results and discussion

As the dataset comprises five semantic categories and twenty-five MDMs, an obvious difficulty lies in the presentation of the results. For the sake of clarity, I will present the results and discussion based on each semantic category. While the columns of *diff. coeff.* and *FS* show the formality scores compared only between the BNC-Acad and the CANCODE corpus, *WFA* takes into account the frequencies from all four corpora. I discuss the formality level of the MDMs in each sub-section on the basis of the comparisons made between the three formality measures. The MDMs shown in each table below are ordered with the most formal at the top and the least formal at the bottom, according to the *WFA* scores. I will discuss to what extent the three calculation methods capture the formality differences and differ from, or resemble, each other in Section 3.8.

3.7.1 Additive MDMs

Seven Additive MDMs were searched: *what's more*, *not only that*, *in addition*, *on top of that*, *besides that*, *add to this*, and *and also*. Table 3.6 shows the observed frequencies (RF) as well as the normalized frequencies (NF) in the four corpora, while Table 3.7 shows the results of the formality scores arrived at by the three formality measures.

Table 3.6

Raw and normalized frequencies of Additive MDMs

MDM	BNC-Acad RF/NF	CANCODE RF/NF	Enronsent2 RF/NF	BASE RF/NF	NF Sum 1
in addition	158/76.17	7/1.44	921/93.73	8/6.74	178.08
add to this	(3)4/1.93	(0)1/0.21	(1)2/0.20	(0)1/0.84	3.18
besides that	1/0.48	1/0.21	3/0.31	1/0.84	1.84
and also	149/71.48	378/77.79	392/39.89	174/146.68	335.84
what's more	3/1.45	12/2.47	11/1.12	4/3.37	8.41
on top of that	2/0.96	12/2.47	12/1.22	3/2.53	7.18
not only that	4/1.93	18/3.70	1/0.10	4/3.37	9.1
NF Sum 2	154.4	88.29	136.57	164.37	

Table 3.7

Formality scores of Additive MDMs

MDM	diff. coeff.	FS	WFA
in addition	-0.96	-0.92	-0.83
add to this	-0.81	-0.60	-0.22
besides that	-0.40	0.00	0.01
and also	0.04	0.43	0.22
what's more	0.26	0.61	0.31
on top of that	0.44	0.71	0.35
not only that	0.32	0.64	0.46

The Sum 1 column and Sum 2 row in Table 3.6 show the general tendencies of Additive MDMs in the four corpora. The most frequently employed Additive MDM in the formal register is *in addition*, whereas it is *and also* in the less formal discourse. Sum 2 shows that Additive MDMs are generally used quite frequently in the four corpora. The fact that the CANCODE corpus has the fewest Additive MDMs might be attributable to the characteristics of English daily conversation, which uses contracted forms and ellipsis to portray the spoken language's spontaneous, rapid nature. Employing a MDM is not as rapid as a SDM.

On the basis of the formality score results, *in addition* is undoubtedly the MDM that stands on the formal end of the continuum, whereas *not only that* sits on the informal end. Although some might consider *and also* tautological, this marker is extensively used not only in written but also in spoken discourse, with a slightly higher frequency in academic spoken discourse, with the NF: 146.68. Although *add to this* is a formal MDM based on the results of the three formality measures, the *WFA* score is arrived at by adding the *add one* rule to the zero frequency observed in the CANCODE and BASE corpora. The numbers in parentheses in Table 3.6 indicate the RF, whereas the number next to the brackets shows the smoothing results. This format applies to the rest of the frequency tables in this chapter.

While both the *diff. coeff.* and *FS* methods show formality scores by comparing two corpora (BNC-Acad and CANCODE), the *WFA* method considers the frequencies from the four corpora. Such a difference leads to the less extreme formality scores observed in Table 3.7. For example, while *add to this* has a very formal weight in both

the *diff. coeff.* (-0.81) and *FS* (-0.60) methods, the *WFA* method took into account the frequencies in the Enronsent2 and the BASE corpus and gave this MDM the formality weight -0.22. One reason that both the *diff. coeff.* and *FS* methods gave *add to this* very formal scores is the zero frequency observed in the CANCODE corpus. The interpretation is that this MDM is so formal that it does not appear in the informal spoken corpus. When considering the frequencies of the four corpora, however, we see that *add to this* also has a zero frequency in the BASE corpus—the database that is academic but spoken-oriented. The less formal score (-0.22) that the *WFA* method calculated then suggests that the formality weight of *add to this* is counterbalanced by the zero occurrences observed in the two corpora, CANCODE and BASE. In other words, the *WFA* method, which considers four frequencies in its calculation, shows how the formality weight might also be influenced by the frequencies.

3.7.2 Appositive MDMs

That is to say, in other words, what I'm saying is, and what this means is are the four Appositive MDMs examined. Table 3.8 shows the frequency counts, whereas Table 3.9 shows the formality scores.

Table 3.8

Raw and normalized frequencies of Appositive MDMs

MDM	BNC-Acad RF/NF	CANCODE RF/NF	Enronsent2 RF/NF	BASE RF/NF	NF Sum 1
that is to say	289/139.33	12/2.47	92/9.36	94/79.24	230.40
in other words	146/70.39	159/32.72	131/13.33	26/21.92	138.36
what I'm saying is	1(0)/0.48	6(5)/1.23	3(2)/0.31	11(10)/9.27	11.29
what this means is	1/0.48	11/2.26	13/1.32	37/31.19	35.25
NF Sum 2	210.68	38.68	24.32	141.62	

Table 3.9

Formality scores of Appositive MDMs

MDM	diff. coeff.	FS	WFA
that is to say	-0.97	-0.92	-0.39
in other words	-0.37	0.04	-0.10
what I'm saying is	0.44	0.71	0.54
what this means is	0.65	0.83	0.58

The Sum 1 column shows that *that is to say* and *in other words* are frequently used Appositive MDMs. The use of Appositive MDMs, however, is more noticeable in the two academic-oriented corpora, the BNC-Acad and the BASE corpus, suggesting that writers and speakers in the two corpora might use Appositive MDMs frequently

when the need to explain or to elaborate on difficult concepts arises. A tentative conclusion is that the register difference influences the use of the MDM semantic type.

The formality score results show that while *that is to say* is a formal, Appositive MDM, the marker-like phrases such as *what I'm saying is* and *what this means is* tend to be employed in the spoken register. At the same time, *in other words* occurs in both formal and informal contexts. The WFA formality scores are less extreme when compared with the other two methods, as is the case in the Additive MDMs, providing a counterbalance to the very formal scores of the *diff. coeff.* and *FS* methods. Take *that is to say* for example. The formality scores -0.97 and -0.92 create an extremely formal image for *that is to say*. Despite that, the frequency information shows that it is also used in other types of discourse. The score -0.39 that the WFA method gives to this MDM shows such a tendency.

3.7.3 Causal MDMs

Four Causal MDMs were searched: *as a result*, *because of that*, *as a consequence*, and *for that reason*. Table 3.10 shows the frequency information across the four corpora, whereas Table 3.11 shows the formality scores.

The numbers in Sum 1 suggest that *as a result* is the most frequently used Causal MDM. Sum 2 shows that the use of Causal MDMs mainly occurs in the formal written (BNC-Acad) or formal spoken (BASE) discourse. The use of Causal MDMs in the CANCODE corpus, on the other hand, is quite limited.

Table 3.10

Raw and normalized frequencies of Causal MDMs

MDM	BNC-Acad RF/NF	CANCODE RF/NF	Enronsent2 RF/NF	BASE RF/NF	NF Sum 1
as a result	90/43.39	14/2.88	256/26.05	19/16.02	88.34
as a consequence	50/24.11	1/0.21	13/1.32	5/4.21	29.85
for that reason	52/25.07	15/3.09	35/3.56	7/5.90	37.68
because of that	15/7.23	26/5.35	24/2.44	12/10.12	25.14
NF Sum 2	99.80	11.53	33.37	36.25	

Table 3.11

Formality scores of Causal MDMs

MDM	<i>diff. coeff.</i>	<i>FS</i>	<i>WFA</i>
as a result	-0.88	-0.73	-0.60
as a consequence	-0.98	-0.96	-0.56
for that reason	-0.78	-0.55	-0.37
because of that	-0.15	0.27	0.14

The formality scores in Table 3.11 indicate that the three formality calculation methods differ from one another. The formality scores of *diff. coeff.* have marked the four MDMs as formal markers, shown by the negative scores. Although *FS* marks *because of that* as an informal MDM (0.27), it also gives a very formal weight to *as a consequence*, as is the case with the *diff. coeff.* method. The *WFA* method, however,

ranks *as a consequence* as the second most formal MDM. The frequency in Sum 1 shows that *as a consequence* occurs rather infrequently when compared with the other formal MDM *as a result*. The WFA has taken the frequency effect into account and therefore the formality score it gives is more reflective of this MDM's actual use.

3.7.4 Concessive MDMs

Seven Concessive MDMs were searched: *despite that*, *having said that*, *but then again*, *in spite of that*, *at the same time*, *on the other hand*, and *but still*. Tables 3.12 and 3.13 show the frequency and formality score results.

Table 3.12

Raw and normalized frequencies of Concessive MDMs

MDM	BNC-Acad RF/NF	CANCODE RF/NF	Enronsent2 RF/NF	BASE RF/NF	NF Sum1
in spite of that	9/4.34	1/0.21	2/0.20	1/0.84	5.95
despite that	10/4.82	4/0.82	4/0.41	2/1.69	7.74
on the other hand	94/45.32	43/8.85	35/3.56	21/17.70	75.43
at the same time	51/24.59	53/10.91	27/2.75	17/14.33	52.58
but still	23/11.09	69/14.20	74/7.53	13/10.96	43.78
having said that	1/0.48	69/14.20	52/5.29	23/19.39	39.36
but then again	28/13.50	944/194.26	81/8.24	117/98.63	314.63
NF Sum2	104.14	243.45	27.98	163.54	

Table 3.13

Formality scores of Concessive MDMs

MDM	<i>diff. coeff.</i>	<i>FS</i>	<i>WFA</i>
in spite of that	-0.91	-0.80	-0.36
despite that	-0.71	-0.43	-0.18
on the other hand	-0.67	-0.37	-0.13
at the same time	-0.39	0.02	0.04
but still	0.12	0.50	0.13
having said that	0.93	0.97	0.69
but then again	0.87	0.94	0.74

Sum 2 in Table 3.12 shows that the use of Concessive MDMs is more noticeable in the two spoken corpora, the CANCODE and the BASE corpus, resulting mainly from the high frequencies of the informal MDM *but then again*. Such tendency is supported by the very high Sum frequency of *but then again* in Sum 1. Enronsent2 uses the fewest Concessive MDMs.

In terms of the formality scores, while the *diff. coeff.* method gives the first four MDMs negative scores, indicating the formal tendency, the *FS* and *WFA* methods mark *at the same time* a neutral MDM, with the other three MDMs marked as formal.

Considering the frequency distribution of *at the same time* across the four corpora, it seems that this MDM functions more like a neutral MDM, rather than a formal marker.

Another difference observed between the *WFA* and the other two methods is seen in their scores of the informal MDMs *having said that* and *but then again*. While the *diff.*

coeff. and *FS* methods give a very informal weight to *having said that*, the *WFA* method marks this MDM as second to the most informal. The frequencies in Sum 1 show that *but then again* occurs nearly eight times more than *having said that* across the four corpora. Such a high frequency effect influences the calculation of the formality score in the *WFA* method. *But then again* is therefore given a more informal weight than *having said that*.

The same frequency leverage is seen in the very formal MDMs, *in spite of that* and *despite that*. Both are given a very high formal score by the *diff. coeff.* and *FS* methods. Although both are considered very formal MDMs, they occur with relatively low frequencies across the four corpora. The *WFA* method therefore gives the two MDMs lower formal scores. Despite the low frequencies observed, *despite that* and *in spite of that* occur mainly in formal, academic writing, whereas *but then again* and *having said that* appear more regularly in a spoken, less formal register. *At the same time* and *on the other hand* tend to be neutral on the continuum.

3.7.5 Contrastive MDMs

Three Contrastive MDMs were searched: *on the other hand*, *on the contrary*, and *by contrast*. The observed frequency results are shown in Table 3.14. Table 3.15 summarizes the formality scores.

The information in Sums 1 and 2 shows that *on the other hand* is the most frequently employed Contrastive MDM across the four corpora. In addition, the Sum 2 column reflects that Contrastive MDMs are found to occur mainly in the BNC-Acad and

the BASE corpus. In the CANCODE and the Enronsent2 corpus, however, Contrastive MDMs are rarely seen.

Table 3.14

Raw and normalized frequencies of Contrastive MDMs

	BNC-Acad	CANCODE	Enronsent2	BASE	NF
MDM	RF/NF	RF/NF	RF/NF	RF/NF	Sum1
by contrast	72/34.71	8/1.65	27/2.75	10/8.43	47.54
on the contrary	40/19.28	2/0.41	5/0.51	10/8.43	28.63
on the other hand	112/54.00	15/3.09	32/3.26	34/28.66	89.01
NF Sum2	107.99	5.15	6.52	45.52	

Table 3.15

Formality scores of Contrastive MDMs

MDM	diff. coeff.	FS	WFA
by contrast	-0.91	-0.80	-0.41
on the contrary	-0.96	-0.91	-0.30
on the other hand	-0.89	-0.76	-0.26

Unlike other semantic categories that have both formal and informal MDMs, all of the three Contrastive MDMs appear to carry more formal weight, shown by the negative formality scores. As is the case in the previous semantic categories, the *WFA* method yields less extreme formality scores. This is particularly evident in *on the other*

hand. The frequencies in Table 3.14 show that *on the other hand* is used in both written and spoken discourse, as well as in the formal and less formal registers. While both the *diff. coeff.* and *FS* methods are indicative of the formality difference, the *WFA* method shows the additional frequency effect.

The CANCODE corpus is found to have fewest Contrastive MDMs, suggesting that speakers do not favour MDMs in showing Contrastive coherence in spoken discourse. The finding that all the Contrastive MDMs were found to occur in a more formal context (BNC-Acad and BASE) might be explained by one of Altenberg's (1986) findings on Contrastive DMs. Altenberg reported that "the type-token ratio of the contrastive links is higher in the written than in the spoken corpus" (1986, p. 26). He observed that when making a Contrastive link in the spoken register, speakers relied heavily on the single-word marker *but*. Writers, on the other hand, "tend to vary their means of expression, lexically and grammatically" (Altenberg, 1986, p. 26). The finding with regard to the Contrastive MDMs from this thesis, and Altenberg's finding on *but*, might serve as supplementary evidence for each case in supporting and explaining the results.

3.8 The comparisons of *WFA*, *diff. coeff.*, and *FS* methods

The *WFA* method considers information gathered from four corpora, whereas the *diff. coeff.* and the *FS* methods only include frequencies from the BNC-Acad and the CANCODE corpus. This difference leads to the formality discrepancies each method assigns to the same set of MDMs. The formality given via the *WFA* approach tends to be less rigid, as the method takes into account the linguistic data from other informal

written or formal spoken databases, increasing the frequency leverage on the formality weight.

From the formality scores shown in the previous tables, a general observation is that the three measures display similar patterns in terms of the formality weight each gives to the target MDMs. Nevertheless, the three methods give slightly different scores to some of the MDMs. Although the *FS* scores are estimated in a similar manner to the *diff. coeff.* calculation (i.e., considering the frequencies from the BNC-Acad corpus and the CANCODE corpus), they differ from one another in some cases. In what follows, I will use *because of that*, *add to this*, and *having said that* to show the advantages and disadvantages of the three methods.

Intuitively, *because of that* seems to be used more frequently in a spoken register; however, the low frequencies of this marker in the BNC-Acad corpus and the CANCODE corpus result in a negative *diff. coeff.*, -0.15. This is counter-intuitive. The *FS* method also considers the frequency information from the BNC-Acad and CANCODE corpora. In this case, it gives a score of 0.27 to *because of that*, indicating that the *FS* method might be the more sophisticated calculation of the two. Nevertheless, including the frequencies in the Enronsent2 corpus and the formal spoken BASE corpus in the calculation of the *WFA* score, I have demonstrated that the formality score can be changed from -0.15 to 0.14, which assigns *because of that* to the informal register. There are three other MDMs that are found to be in the same category where only the *diff. coeff.* method assigns formal weight, but not the *FS* nor the *WFA* method: *at the same time*, *because of that*, and *besides that*.

Add to this, according to the *diff. coeff.* and the *FS* method, is a very formal Additive MDM, with formality scores of -0.81 and -0.60 respectively. When looking at the distributional patterns of *add to this* in the four corpora in Table 3.6, one cannot help but notice the sparseness of this marker in the four corpora. The very formal weight assigned by the above two methods suggests this marker's importance in the written register. Such inference, however, is misleading, as *add to this* only occurs 1.93 times (NF) in the BNC-Acad and 0.21 times in the CANCODE corpus. The score of -0.40 implies that a higher frequency count is to be found in the BNC-Acad corpus, and yet this is not the case. The frequencies serve as counter-evidence in relation to the formality weight given to *add to this*, -0.80 and -0.60. The *WFA* method, on the other hand, assigns this marker a weight of -0.22 on the basis of its distribution across the four corpora. I propose that extra care has to be taken when interpreting the formality scores arrived from two corpora only. In other words, the *WFA* method shows the influence that the low frequency of *add to this* has upon its formality weight. Two other MDMs that are found in the same circumstance where both the *diff. coeff.* and *FS* methods assign very formal weight, but the *WFA* method does not, are *as a consequence* and *on the contrary*.

A similar phenomenon in which the low frequency affects the assigned formality weight is observed among the informal MDMs. Both the *diff. coeff.* and *FS* methods assign a very informal weight to the Concessive MDM *having said that*, with 0.93 from the former, and 0.97 from the latter. The NF of this marker in the CANCODE and BNC-Acad corpora are 14.20 and 0.48 respectively.

Comparing the frequencies of *having said that* with the frequencies of *but then again* in the two corpora, which are 194.26 in the CANCODE corpus and 13.50 in the BNC-Acad corpus, we see that the latter MDM occurs relatively frequently in the informal register. The *WFA* method considers the frequencies from the four corpora and assigns 0.67 to *having said that* and 0.74 to *but then again*. Another MDM that is found to be in the same scenario, where both the *diff. coeff.* and *FS* methods assign very informal weight but the *WFA* method does not, is *on top of that*.

The above discussion has supported the rationale of including frequencies from other corpora in calculating an MDM's formality score. Looking at the formality scores assigned by either the *diff. coeff.* or the *FS* method, one has a clear indication of how formal or informal a MDM is. However, a very formal or informal score assigned to one MDM with these two methods does not reflect how frequently this MDM is used. One clear contribution that the *WFA* method makes to the two existing methods is that it includes frequency information from varied corpora to leverage the calculation.

3.9 The formality continuum

From the frequency as well as the formality scores shown in this chapter, a formality continuum is created (see Figure 3.2). The corpus evidence suggests that the formality of the context has an influence on what MDMs language users choose to employ. Within each semantic category, each MDM can be allocated a place in terms of the formality level.

The formality continuum of MDMs

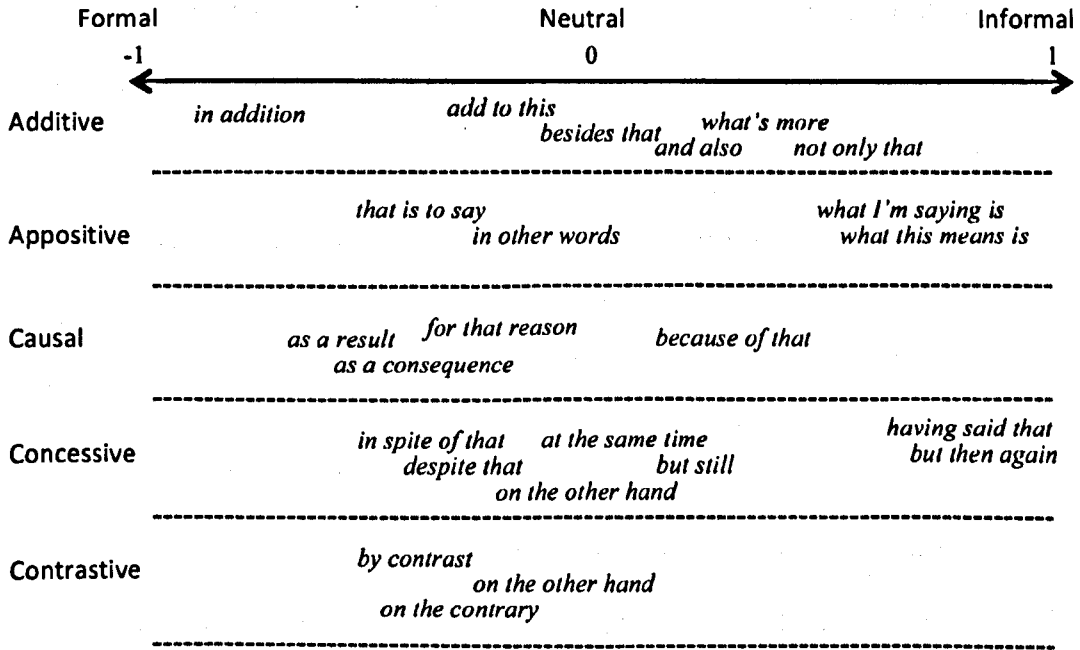


Figure 3.2 The formality continuum of MDMs

From the formality continuum above, I propose an alternative classification of discourse markers, which are conventionally classified via the coherence relations they signal. Within each coherence relation, a further subtle categorization—the formality difference—can be used to account for the difference between, for example, *despite that* and *having said that*. In other words, I have demonstrated how to perceive the differences and patterns of synonymous MDMs by taking into account textual formality. This finding has provided a clear answer for the first research question that I raised in Chapter 1:

RQ1. Does a formality continuum exist within each semantic relation that distinguishes synonymous MDMs, for example, differentiating *what's more* from *in addition* within the Additive semantic category?

The answer to this question is yes, as is evidenced by the continuum shown in Figure 3.2.

3.10 Summary and conclusion

In this chapter, I have attempted to describe the formality patterns of MDMs in both spoken and written English as well as in formal and informal contexts. The devised *WFA* method, which considers the dimension of difference factors, has proven to be heuristically sound when compared with existing formality calculation methods. Not only have I identified the most frequently used MDMs exclusive to spoken and written registers, I have also pointed out some neutral MDMs that can be used in both the spoken and the written register, supporting the dynamic nature of language in use.

I have only examined NS data in this chapter. In Chapter 4, I will investigate how L2 learners employ MDMs in their language output. Whether or not L2 learners exhibit a pattern similar to that of NS is one potentially interesting avenue of exploration. When L2 learners tend to prefer one type of MDM in an atypical context (for example, using more formal MDMs in an informal context or vice versa), such behaviour is conventionally perceived as an error or misuse; yet such misuse has its pedagogical and learning implications. The observed pattern of MDMs on the formality continuum

will be used as a reference when I explore L2 learners' pragmatic awareness in their use of MDMs.

Chapter 4 - Discourse markers in use: evidence from the Cambridge Learner Corpus

On the basis of the corpus evidence drawn from the four reference corpora, and using the Weighted Formality Average method (WFA) to calculate the formality weight of multi-word discourse markers (MDMs), I have shown in Chapter 3 that a formality continuum exists among various MDMs. The formality continuum gives a clear account of the context in which language users prefer a certain group of MDMs to others. For example, in terms of Causal MDMs, *as a consequence* and *as a result* are employed mainly in the formal register whereas *because of that* is used more frequently in the informal, spoken register. The only exception is the Contrastive semantic category. The examined Contrastive MDMs are found to be either formal or neutral in terms of the formality continuum. No informal Contrastive MDMs are observed.

The literature reported in Chapter 2 indicates that L2 learners show a lack of sensitivity to register difference when they use discourse markers (DMs) in writing. In most of the L2 DMs studies, the unit of analysis is mainly single-word discourse markers (SDMs). The inclusion of MDMs in those studies was limited to those that are frequently used by L2 writers, such as *on the other hand*. In other words, questions such as to what extent L2 writers use MDMs in creating semantic links, and further, whether or not L2 writers show a lack of sensitivity to registers in using MDMs, remain unexplored. In this chapter, I will examine L2 learners' use of MDMs in the Cambridge Learner Corpus (CLC). The goal is to provide a detailed description of the distributional pattern of the MDMs that L2 learners employ. The L2 MDMs usage pattern will be compared with the data reported in Chapter 3.

The learner corpus data will be examined qualitatively and quantitatively. The qualitative approaches taken in examining learner data are similar to those used for NS data, yet different in a few aspects. The areas of similarity include the manual examination, which is required due to the multi-functional and polysemous features of MDMs (see Section 3.6.1). Looking at learner data, however, requires further consideration.

MDMs are conventionally treated as a subset of formulaic language that shows the frozen phrase nature (Wray, 2002). The fixed nature of MDMs' constituents, however, does not suggest that that acquisition of MDMs is an all-or-nothing phenomenon. Fixed or conventional phrases are usually comprised of both function words (prepositions, such as *in*, *on*, and *of*) and content words (nouns, verbs, adjectives, etc.). Before learners can fully grasp the usage of multi-word units, it is possible that a function word within a fixed phrase will be replaced by another function word. Take the Additive MDM *in addition* and its other form *in addition to that/this* for example. A learner who uses *in addition to that* is considered to have made a successful attempt at using the Additive MDM *in addition*; yet, the exploration of the CLC learner data shows that at times, the function word *to* is replaced by another function word. An attested learner sentence illustrates such an unsuccessful attempt, shown in sentence (1).

(1) *In addition of this*, I grew up with two brother younger than me who gave me
a good feeling with children. (C2, 568085_1)

When searching MDMs in the CLC corpus, I exclude examples like the above, as the marker used is only partially correct.

Another commonly observed phenomenon when going through learner data is the misuse of MDMs. 'Misuse' here refers to the confusion of meaning, rather than to syntactic mistakes. Another attested learner example illustrates this, shown in sentence (2).

(2) I love camping and I am used to staying outside. *On the other hand*, I am good at any kind of sports so I think I will be able to give my hand when you have some sports activities. (B2, 565482_2)

In his job application letter, this B2 learner has mentioned his competence in English, French and his major in international tourism. In his attempt to create an Additive link so that he could mention his skill at sports as an additional advantage, this B2 learner used *on the other hand* as the linking marker, instead of a more appropriate marker such as *in addition* or *furthermore*. I disregard misuse cases (caused by meaning confusion) from the data set. In other words, by filtering out misuse cases via manual examination, the collected data will provide a more reliable sample for analysis, which better represents the use patterns of MDMs in the CLC corpus.

Another aspect that deserves some attention is when learners use the target MDMs correctly. The correct uses of MDMS by NS and by L2 learners are different in

the sense that the former usually accompanies correct grammar and wording in the remainder of the sentences. The correct use of MDMs by learners, on the other hand, usually co-occurs not only with correct but also with problematic grammar or wording in the preceding or following sentences. As the targeted items are MDMs, I therefore include cases where MDMs are found to co-occur with ungrammatical structures or awkward wording.

The quantitative approach taken in analysing learner data includes normalized frequency (NF), the log-likelihood (*LL*) test, and Pearson's *r* (Rodgers & Nicewander, 1988; Stigler, 1989). McEnery, Xiao, & Tono (2006) introduced various tests to show statistical significance in corpus-based studies, including the log-likelihood (*LL*) score. The *LL* test is used to see whether or not the observed distributional pattern of an MDM in one sub-corpus is different from the pattern of the same MDM in another sub-corpus. In other words, I use NF and *LL* tests to see how learners' English proficiency affects their use of MDMs in two dimensions: a vertical and a horizontal one. The vertical dimension will show progression in the quantity of MDMs used through the levels of learners. The horizontal dimension will show how the different MDMs in each semantic category are used across the formality continuum. Pearson's *r* will be used to investigate the relationship between the formal and informal MDMs.

Like the data in the NS corpora, I do not consider the position of MDMs in the analysis. The literature reviewed in Chapter 2 suggests that DMs often occur at the initial and median position of a discourse unit, regardless of whether the unit refers to a sentence, an utterance, or a turn. Various positions of the same discourse marker might suggest a change or a continuity of a discourse. As the focus of this research lies

in L2 learners' sociolinguistic competence in using MDMs, i.e., the formality consideration when employing various markers, I consider the syntactic position of MDMs less relevant and it is therefore left out of the account.

4.1 The Cambridge Learner Corpus (CLC)

The Cambridge Learner Corpus (CLC) is a large collection of learner English in written format, with 45 million words at the time I conducted the research. Despite its substantial collection of learner data, CLC is only accessible to the English Profile research teams. The English Profile aims to create a profile of "the English language levels of learners in terms of the six proficiency bands of the CEFR, A1 to C2" (Using the CEFR, 2011, p.31). The CEFR (Common European Framework of Reference) is a framework that "describes language learners' ability in terms of speaking, reading, listening, and writing at six reference levels" (Using the CEFR, 2011, p.4). Table 4.1 shows a summary of the six levels. Following the successful submission of my research proposal to the English Profile research project, I was granted access to the corpus online via Sketch Engine.

The main data archived in the CLC corpus are the Cambridge ESOL English exam scripts completed by English learners around the world. The metadata of the exam scripts include L1 background, age, IELTS levels, CEFR reference levels, the register of the writing, etc. On the basis of the writing style and the nature of the text, every exam script is labelled as formal, informal, neutral, or mixed under the category *register*. Learner writing in the CLC formal register includes argumentative essays, compositions, and reports in which learners express their opinions on controversial

issues. Writing in the informal register, on the other hand, includes writing that is classified as descriptive or creative autobiographic type writing, such as letters of greeting, and advice to friends. As the notion of register is the main theme in this study and one of the goals is to explore how learners' English level affects the use of MDMs, I will examine the CLC data from two angles: the formality of the writing, and the CEFR levels.

Table 4.1

The six levels of the CEFR

Band	Description	Level
C2	Mastery	Proficient user
C1	Effective operational proficiency	
B2	Vantage	Independent user
B1	Threshold	
A2	Waystage	Basic user
A1	Breakthrough	

The CLC corpus on Sketch Engine enables users to create a sub-corpus on the basis of a number of criteria. I specifically set the two metadata, register (formal and informal) and CEFR bands as the criteria and created seven sub-corpora, shown in Table 4.2. Note that the current CLC database does not include any informal writing produced by C2 learners, nor formal and informal writing produced by A learners. The lack of C2 informal writing makes the comparison between formal and informal

writing at the C2 level less straightforward. Nevertheless, comparisons between the C1 levels are still plausible. The learner levels compared are therefore mainly between levels C and B: that is, between advanced and intermediate learners.

Table 4.2

Summaries of the seven sub-corpora in CLC

Register	CEFR level	Corpus name	Size (words)
Formal	C2	F-C2	1,238,165
	C1	F-C1	1,734,691
	B2	F-B2	1,612,329
	B1	F-B1	454,971
Informal	C1	Inf-C1	101,099
	B2	Inf-B2	568,665
	B1	Inf-B1	1,099,519

The presentation of the data and results is structured as follows. First, I will present the data in accordance with the semantic type of MDMs. Within each semantic category, I will show data obtained from the four formal sub-corpora, followed by the data from the three informal sub-corpora. Analysis and comparison of the two types of data will be made within each semantic category.

4.2 CLC data in the formal and the informal register

All the targeted MDMs listed in Chapter 3 were searched in the four formal and the three informal sub-corpora. Discussion in each semantic category will confine itself to the following issues: the observed frequency, the comparisons of the normalized frequency (NF) among the seven sub-corpora, and the formal-informal ratios of MDMs employed in each register.

4.2.1 Additive MDMs

4.2.1.1 Formal register

There are seven types of MDM in this category: *what's more, in addition, on top of that, besides this/that, and also*, and *add to this*. Some of the Additive MDMs have their sub-forms. For example, *in addition to this/that* is classified as one sub-type of *in addition*; likewise *and on top of that* is counted as one instance of *on top of that*. Table 4.3 shows the raw frequency (RF) and the NF of these MDMs in the four formal sub-corpora. Markers in Table 4.3 are placed in accordance with the formality weight found in Chapter 3; that is, instead of a horizontal formality continuum, the first column represents a vertical formality continuum, with more formal markers at the top and more informal markers at the bottom. This mechanism also applies to other tables throughout Chapter 4.

The data above show that C1 and B2 learners used the most Additive MDMs, C2 learners showed a moderate approach, whereas B1 learners used the least. The top two Additive MDMs produced across the four groups of learners are the formal MDM *in addition*, and the informal MDM *and also*.

Table 4.3

RF and NF of Additive MDMs in CLC, formal register

MDMs	Raw frequency				NF (per million words)			
	F-C2	F-C1	F-B2	F-B1	F-C2	F-C1	F-B2	F-B1
in addition	382	747	658	33	308.52	430.62	408.11	72.53
add to this	6	5	6	0	4.85	2.88	3.72	0.00
beside this/that	28	52	47	2	22.61	29.98	29.15	4.40
and also	237	495	504	180	191.41	285.35	312.59	395.63
what's more	33	52	32	0	26.65	29.98	19.85	0.00
on top of that	15	45	11	0	12.11	25.94	6.82	0.00
not only that	9	11	4	1	7.27	6.34	2.48	2.20
Sum					573.43	811.10	782.72	474.76

I ran *LL* tests to see whether the differences among the four corpora are statistically significant. Paired comparisons were made between C2 and C1, C1 and B2, and finally, between B2 and B1. The reason for not comparing, for example, C2 data with B1 data is simply because the two learner groups differ substantially in terms of English proficiency and such comparison might be less meaningful. The *LL* test results of the three pairs are shown in Table 4.4.

Table 4.4

LL test results of Additive MDMs, formal register

MDMs	C2 vs. C1		C1 vs. B2		B2/B1	
	LL ratio	Sig. level	LL ratio	Sig. level	LL ratio	Sig. level
in addition	29.003	.000** (C1)	1.010	0.365, ns	161.832	.000** (B2)
add to this	0.739	0.390, ns	0.179	0.672, ns		NA
beside this/that	1.483	0.223, ns	0.019	0.890, ns	12.708	.000** (B2)
and also	26.624	.000** (C1)	2.075	0.150, ns	7.080	0.008 (B1)
what's more	0.281	0.596, ns	3.457	0.063, ns		NA
on top of that	7.279	0.007, (C1)	19.734	.000** (C1)		NA
not only that	0.092	0.792, ns	2.905	0.088, ns	0.012	0.913, ns

** p< .001 level

When a zero frequency occurs, which makes it virtually impossible to calculate *LL* scores, I use NA (not applicable) to indicate such a case. When the *LL* test result reaches the significance level, I indicate where the difference lies by putting the corpus code in parentheses in the Sig. level column. For example, the result of the first pair comparison (C2 vs. C1) in the formal MDM *in addition* showed that the use of this marker by C2 learners was very different from the pattern observed in C1 learners: shown statistically, $LL = 29.003$, $p < 0.001$. The difference lay in the fact that C1 learners showed a higher frequency of employing *in addition*, $NF = 430.62$. I marked the predominance of frequency in the Sig. level column by putting the corpus code,

shown as (C1). In comparison with C2 learners, C1 learners also preferred two informal, Additive MDMs, *and also* and *on top of that*.

In the second pair comparison (C1 vs. B2), the results showed that the two learner groups used Additive MDMs in a similar manner, as the *LL* tests indicated that only one MDM (*on top of that*) was found to be significantly different. The third pair comparison (B2 vs. B1) yielded some interesting results. The observed frequency data in Table 4.3 showed that B1 learners used neither *add to this*, *on top of that* nor *what's more*. B1 learners mainly employed two types of Additive MDM, the formal marker *in addition* and the informal *and also*, suggesting that B1 learners used rather limited linguistic devices in this regard. Such dearth of other Additive MDMs from B1 learners might explain their over-reliance on the informal marker *and also*, as this is the only marker in which B1 learners showed a higher frequency when compared with B2 learners.

When grouping the seven MDMs on the basis of formality weight, we have three groups: the formal group (2 MDMs: *in addition* and *add to this*), the neutral group (1 MDM: *besides this*), and the informal group (4 MDMs: *and also*, *what's more*, *on top of that*, and *not only that*). That is, by converting the NF into percentages, we could compare the formal-informal ratio of MDMs in the four groups of learners. For example, C2 learners produced 313.37 instances of formal MDMs (*in addition*: 308.52 + *add to this*: 4.85) out of the total 573.43 MDMs found. The percentage of formal MDMs in the F-C2 corpus would be 55 percent, arrived at from the following formula:

$$\left(\frac{308.52}{573.43}\right) \times 100$$

This means that more than half (55%) of the Additive MDMs that F-C2 produced were formal-oriented MDMs. Applying the same mechanism to the NF of informal MDMs, we would then have a clear picture of the formal-informal MDM ratios. The neutral MDM, *besides this/that*, was not considered in this ratio comparison because of its low frequency. Figure 4.1 shows the formal-informal ratios of the Additive MDMs in the four sub-corpora, described as percentages.

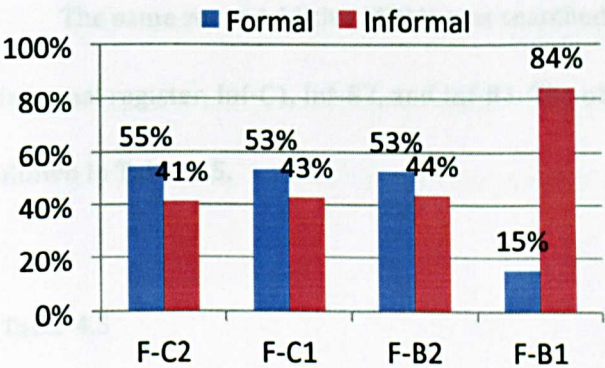


Figure 4.1 Percentages of formal and informal Additive MDMs, formal register

The formal-informal ratios in the F-C2, the F-C1, and the F-B2 corpus showed a similar pattern; that is, the majority of the total MDMs were formal MDMs, whereas less than half were informal. There is, however, a gradual decrease in the rates of formal MDMs as we move from the F-C2 corpus to the F-B1 corpus, and at the same time, a progressive increase in the percentage of informal MDMs. Such a decrease/increase phenomenon would be more marked if we were to include the ratio of the F-B1 corpus. A Pearson product-moment correlation coefficient was computed to assess the relationship between the ratios of the formal and informal Additive MDMs. There was a negative correlation between the two types of MDM ($r = -1.000$, n

= 4, $p = 0.000$). This means that a linear correlation relationship exists between learners' use of the formal and informal MDMs. Overall, increases in the frequency of the formal MDMs were found to correlate with decreases in the frequency of the informal MDMs. The formal-informal ratio data will be compared and discussed with those found in the informal register.

4.2.1.2 Informal register

The same set of Additive MDMs was searched in the three sub-corpora in the informal register, Inf-C1, Inf-B2, and Inf-B1. The observed frequencies and NF are shown in Table 4.5.

Table 4.5

RF and NF of Additive MDMs in CLC, informal register

MDMs	Raw frequency			NF (per million words)		
	Inf-C1	Inf-B2	Inf-B1	Inf-C1	Inf-B2	Inf-B1
in addition	32	88	18	316.52	154.75	16.37
add to this	0	2	0	0.00	3.52	0.00
beside this/that	2	24	3	19.78	42.20	2.73
and also	20	191	218	197.83	335.87	198.27
what's more	16	54	3	158.26	94.96	2.73
on top of that	3	10	1	29.67	17.59	0.91
not only that	0	1	4	0.00	1.76	3.64
Sum				722.06	650.65	224.64

Despite the informal register labelled with the three sub-corpora, both the C1 and B2 learners produced many instances of the very formal MDM, *in addition*; yet, these two groups of learners also used many instances of informal MDMs, such as *and also* and *what's more*. A declining trend in the total number of MDMs (NF) used by the three groups was observed, with C1 learners the highest, B2 learners in the middle, and B1 learners the lowest.

As there are only three sub-corpora in the informal register, the *LL* tests were carried out only for the comparisons made between C1 and B2 learners, and between B2 and B1 learners. Table 4.6 gives a summary of the *LL* test results.

Table 4.6

LL test results of Additive MDMs, informal register

MDMs	C1 vs. B2		B2 vs. B1	
	<i>LL</i> ratio	Sig. level	<i>LL</i> ratio	Sig. level
in addition	10.63	0.001 (C1)	107.827	.000** (B2)
add to this	NA		NA	
beside this/that	1.316	0.251, ns	35.321	.000** (B2)
and also	5.853	0.016 (B2)	27.643	.000** (B2)
what's more	2.922	0.087, ns	95.222	.000** (B2)
on top of that	0.572	0.449, ns	15.656	.000** (B2)
not only that	NA		0.483	0.487, ns

** $p < .001$ level

C1 learners and B2 learners displayed a similar usage pattern in informal MDMs, as shown by the non-significant findings between the two groups' uses of *what's more* and *on top of that*. Nevertheless, C1 learners used more instances of the formal MDM *in addition*, whereas B2 learners produced more examples of *and also*, the informal MDM. In the second pair comparison, B2 learners produced more Additive MDMs in general, suggesting B2 learners' predominant linguistic ability in producing both formal and informal MDMs.

Converting the NF to percentages showed a clearer picture of the formal-informal ratios as well as the distributions across the three sub-corpora (see Figure 4.2).

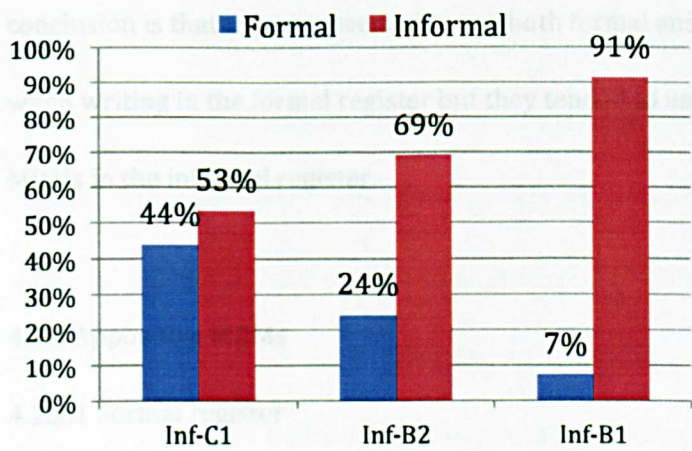


Figure 4.2 Percentages of formal and informal Additive MDMs, informal register

The red bars (informal MDMs) stood in marked contrast to the blue bars (formal MDMs) which seemed rather straightforward, as it was the informal type of writing that was examined in this section; however, comparing the differences observed

between the blue and red bars in Figures 4.1 and 4.2, we see how the L2 learners used the formal and informal Additive MDMs in the two registers.

In the formal register, we see that the differences between the blue and red bars in the four sub-corpora are 14% in the F-C2 corpus, 10% in the F-C1 corpus, 9% in the F-B2 corpus, and 69% in the F-B1 corpus respectively (see Figure 4.1). In Figure 4.2, the gaps between the blue and red bars were widened: this was particularly noticeable in the Inf-B2 corpus (45%), and in the Inf-B1 corpus (84%). The two groups of learners' increased use of informal MDMs and their decreased use of formal MDMs in the informal register gave rise to such widening. Although the lack of data from the Inf-C2 corpus makes the direct comparisons of the two figures less straightforward, by looking at the C1, B2 and B1 data in both the formal and the informal register, one conclusion is that learners seemed to use both formal and informal Additive MDMs when writing in the formal register but they tended to use more informal Additive MDMs in the informal register.

4.2.2 Appositive MDMs

4.2.2.1 Formal register

The Appositive MDM category includes four types of MDM: *that is to say, in other words, what I'm saying is*, and *what this/that means is*. Other sub-types of Appositive MDMs include *what I mean is* and *that is*. Table 4.7 shows the observed raw frequency (RF) and NF of these MDMs in the four formal sub-corpora.

Table 4.7

RF and NF of Appositive MDMs in the four learner corpora, formal register

MDMs	Raw frequency				NF (per million words)			
	F-C2	F-C1	F-B2	F-B1	F-C2	F-C1	F-B2	F-B1
that is to say	68	23	21	4	54.92	13.26	13.02	8.79
in other words	96	15	19	1	77.53	8.65	11.78	2.20
what I'm saying is	1	0	0	0	0.81	0.00	0.00	0.00
what this means is	0	0	0	0	0.00	0.00	0.00	0.00
Sum					133.26	21.91	24.81	10.99

An obvious observation from Table 4.7 is that other than C2 learners, the other three groups of learners used relatively few instances of Appositive MDMs. As was the case with Additive MDMs, B1 learners used the least Appositive MDMs. The fact that C1, B2, and B1 learners produced a relatively small number of Appositive MDMs might be explained by the more sophisticated skills required in re-working or reformulating arguments (Hyland, 2007). By contrast, as C2 learners presumably have a richer linguistic repertoire, they might be more confident in reworking their arguments and therefore giving rise to a much higher frequency of use of Appositive MDMs.

As only one instance of informal MDMs occurs in the four sub-corpora, *LL* tests were carried out in the comparisons of the two MDMs, *that is to say* and *in other words*. The *LL* test results are shown in Table 4.8.

Table 4.8

LL test results of Appositive MDMs, formal register

MDMs	C2 vs. C1		C1 vs. B2		B2/B1	
	LL ratio	Sig. level	LL ratio	Sig. level	LL ratio	Sig. level
that is to say	41.009	.000** (C2)	0.003	0.953, ns	0.566	0.452, ns
in other words	96.409	.000** (C2)	0.810	0.368, ns	4.532	0.033 (B2)

** p< .001 level

The first pair comparison (C2 vs. C1) showed that C2 learners used far more formal and the neutral MDMs than did C1 learners. The *LL* tests in the second pair comparison (C1 vs. B2), however, showed no significant differences, suggesting that the use of Appositive MDMs by C1 learners was similar to B2 learners. Such a similarity was also observed in the Additive category. In the third pair, where I compared B2 with B1 learners, the only significant difference lay in B2 learners' frequent use of the neutral MDM, *in other words*.

In terms of formality, the majority of learners in the four sub-corpora tended to use two markers only: the formal marker *that is to say*, and the neutral MDM *in other words*. The formal-informal ratio comparison was therefore made between the formal and the neutral MDM, shown in Figure 4.3.

The formal-neutral ratio showed that while C1, B2, and B1 learners produced a greater percentage of *that is to say*, C2 learners used more neutral MDM *in other words*.

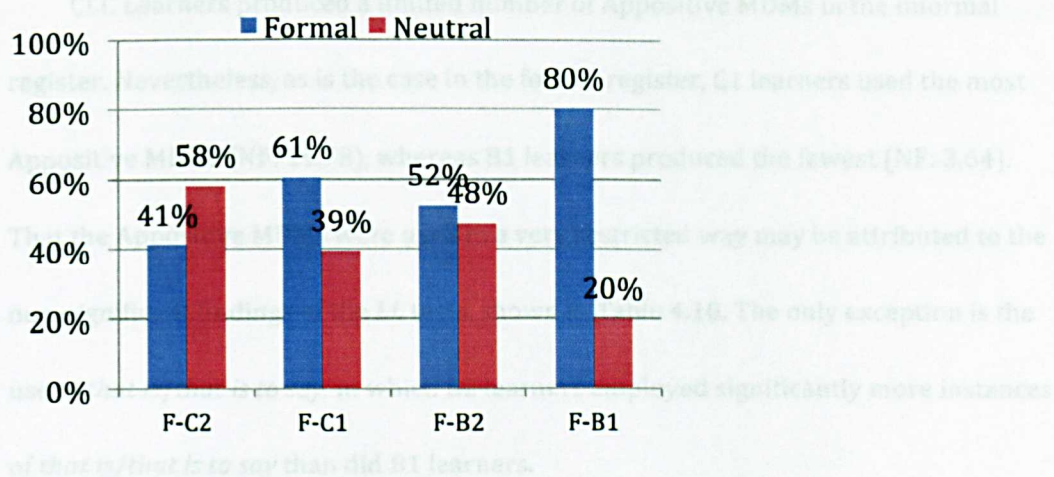


Figure 4.3 Percentages of formal and neutral appositive MDMs, formal register

Table 4.10

4.2.2.2 Informal register

The same set of Appositive MDMs was searched in the three sub-corpora in the informal register. The results are shown in Table 4.9.

Table 4.9

RF and NF of Appositive MDMs in the three learner sub-corpora, informal register

MDMs	Raw frequency			NF (per million words)		
	Inf-C1	Inf-B2	Inf-B1	Inf-C1	Inf-B2	Inf-B1
that is to say	1	4	1	9.89	7.03	0.91
in other words	1	3	2	9.89	5.28	1.82
what I'm saying is	0	0	0	0.00	0.00	0.00
what this means is	0	0	1	0.00	0.00	0.91
Sum				19.78	12.31	3.64

CLC Learners produced a limited number of Appositive MDMs in the informal register. Nevertheless, as is the case in the formal register, C1 learners used the most Appositive MDMs (NF: 19.78), whereas B1 learners produced the fewest (NF: 3.64). That the Appositive MDMs were used in a very restricted way may be attributed to the non-significant findings of the *LL* tests, shown in Table 4.10. The only exception is the use of *that is/that is to say*, in which B2 learners employed significantly more instances of *that is/that is to say* than did B1 learners.

Table 4.10

LL test results of Appositive MDMs, informal register

MDMs	C1 vs. B2		B2 vs. B1	
	<i>LL</i> ratio	Sig. level	<i>LL</i> ratio	Sig. level
that is to say	0.087	0.768, ns	4.439	0.0345 (B2)
in other words	0.265	0.607, ns	1.395	0.238, ns
what I'm saying is	NA		NA	
what this means is	NA		NA	

** p< .001 level

The very limited number of Appositive MDMs made it difficult to claim statistical representation, which inevitably led to a problematic interpretation of the formal-informal ratio. For this reason, the formal-informal ratio comparisons were disregarded in this section.

In view of the fact that few instances of Appositive MDMs were found in both the formal and the informal register (with the exception of the F-C2 corpus), it seems that the encoding discourse strategy of employing Appositive MDMs is still not widely adopted by intermediate to lower level learners. A more detailed discussion in this regard is given in Section 4.3.

4.2.3 Causal MDMs

4.2.3.1 Formal register

Four MDMs are included in this category: *as a result*, *as a consequence*, *because of that/this*, and *for that/this reason*. Other sub-types of the four MDMs include *as a result of this/that*, and *as a consequence of this/that*. Table 4.11 summarizes the frequency information of these MDMs, with the first three MDMs categorised as formal, and the last MDM as informal. Data reported in Chapter 3 indicated that no neutral MDMs were found in the Causal semantic category.

Table 4.11

RF and NF of Causal MDMs in the four learner corpora, formal register

MDMs	Raw frequency				NF (per million words)			
	F-C2	F-C1	F-B2	F-B1	F-C2	F-C1	F-B2	F-B1
as a result	317	283	111	1	256.02	163.14	68.84	2.20
as a consequence	65	44	13	1	52.50	25.36	8.06	2.20
for that reason	99	74	79	23	79.96	42.66	49.00	50.55
because of that	33	55	68	5	26.65	31.71	42.18	10.99
Sum					415.13	262.87	168.08	65.94

Looking at the sum column, I observed a declining trend in the number of MDMs produced from the F-C2 to F-B1 corpus, and this decline paralleled the level of English proficiency of the four groups. C2, C1, and B2 learners preferred *as a result*, whereas B1 learners frequently used *for that reason*. C2 learners' least used MDM was the informal MDM *because of that*. While both C2 and C1 learners made use of the formal MDM *as a consequence*, B2 and B1 learners only used this MDM in a limited manner. A look at the BNC word frequency list (Kilgariff, 1997) reveals that the word *consequence* is placed in the 1257th position. The fact that B2 and B1 learners produced fewer instances of *as a consequence* can be plausibly ascribed to the low frequency of the word *consequence*. This low-frequency effect was less profound in C-level learners. Considering the formality types of MDM in the four sub-corpora, C2 and C1 learners seemed to have developed a greater awareness of register difference.

Despite the gradual decreases in the sum column from the F-C2 corpus to the F-B1 corpus, I ran additional *LL* tests to see if the differences were statistically valid. Table 4.12 shows the *LL* test results for the three pairs.

The first pair comparison showed that C2 and C1 learners produced a similar proportion of the informal Causal MDM *because of that*; yet, C2 learners evidently produced more of the other three formal MDMs than did C1 learners. In the second pair comparison, the differences were found to be in the two groups' use of *as a result* and *as a consequence*. In the cases of *for that reason* and *because of that*, C1 and B2 learners showed similar usage patterns. Both B2 learners and B1 learners used the formal MDMs *for that reason* and *as a consequence* in a similar way; yet, at the same time, B2 learners used more instances of *because of that* and *as a result*.

Table 4.12

LL test results of Causal MDMs, formal register

MDMs	C2 vs. C1		C1 vs. B2		B2/B1	
	LL ratio	Sig. level	LL ratio	Sig. level	LL ratio	Sig. level
as a result	30.361	.000** (C2)	65.612	.000** (C1)	46.779	.000** (B2)
as a consequence	14.236	.000** (C2)	15.617	.000** (C1)	2.285	0.131, ns
for that reason	16.949	.000** (C2)	0.734	0.392, ns	0.017	0.895, ns
because of that	0.630	0.427, ns	2.492	0.114, ns	12.482	.000** (B2)

** p< .001 level

The formal-informal ratios in the four sub-corpora are shown in Figure 4.4.

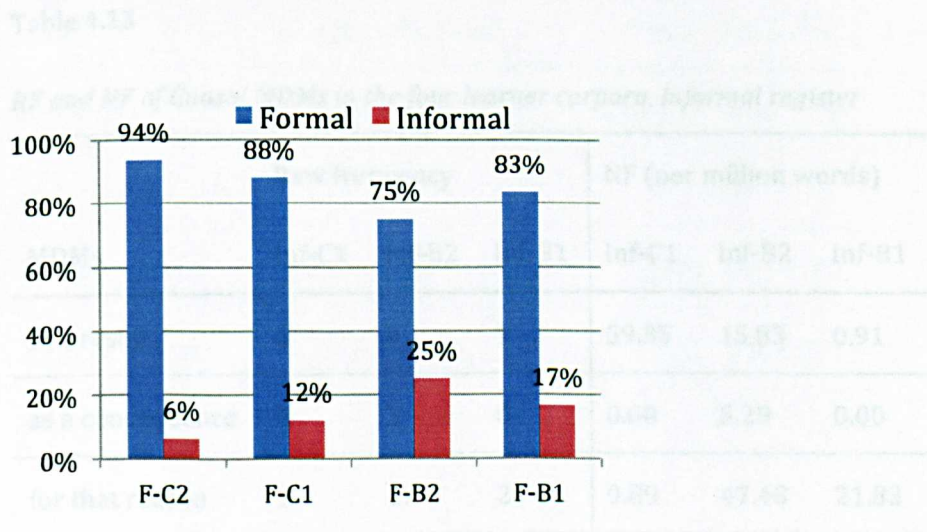


Figure 4.4 Percentages of formal and informal Causal MDMs, formal register

Other than the blue bar in the F-B1 corpus, the proportion of formal MDMs in the other three sub-corpora decreased as the English proficiency level declined; at the same time, the proportion of informal MDMs increased. This is a similar pattern to that

observed with Additive MDMs. Another Pearson product-moment correlation coefficient was computed to assess the increase and decrease. As was the case in Additive MDMs, there was a negative correlation between the two types of MDM ($r = -1.000$, $n = 4$, $p = 0.000$). A tentative conclusion regarding learners' use of Causal MDMs in the formal register is that, as learners' English level developed, the use of formal MDMs increased and the use of informal MDMs decreased.

4.2.3.2 Informal register

I searched the same set of Causal MDMs in the three sub-corpora in the informal register. The observed frequencies and the NF are summarized in Table 4.13.

Table 4.13

RF and NF of Causal MDMs in the four learner corpora, informal register

MDMs	Raw frequency			NF (per million words)		
	Inf-C1	Inf-B2	Inf-B1	Inf-C1	Inf-B2	Inf-B1
as a result	6	9	1	59.35	15.83	0.91
as a consequence	0	3	0	0.00	5.28	0.00
for that reason	1	27	24	9.89	47.48	21.83
because of that	1	20	18	9.89	35.17	16.37
Sum				79.13	103.75	39.11

Since the writing examined here was categorized as informal, it is perhaps not surprising to find that the formal MDM *as a consequence* occurred rarely. The three

groups of learners still produced other formal MDMs, such as *as a result* and *for that reason*; yet, proportionally speaking, the use of the informal MDM *because of that* increased. This is further addressed in a later paragraph when presenting the formal-informal ratio. In terms of the total amount of MDMs produced, B2 learners produced the most Causal MDMs, whereas B1 learners still produced the fewest.

In order to see whether the differences in frequencies were statistically sound, I ran another *LL* test and present the results in Table 4.14.

Table 4.14

LL test results of Causal MDMs, informal register

MDMs	C1 vs. B2		B2 vs. B1	
	<i>LL</i> ratio	Sig. level	<i>LL</i> ratio	Sig. level
as a result	5.445	0.020 (C1)	13.704	.000** (B2)
as a consequence	NA		NA	
for that reason	3.989	0.046 (B2)	7.599	0.006 (B2)
because of that	2.286	0.131, ns	5.481	0.019 (B2)

** $p < .001$ level

C1 and B2 learners showed a similar usage pattern in their use of the informal MDM *because of that*. When comparing B2 with B1 learners, we see that B2 learners obviously employed more formal and informal MDMs than did B1 learners. This is a similar finding to the *LL* tests in informal Additive MDMs where B2 learners produced

far more instances of both formal and informal MDMs. Such phenomena might be attributed to B1 learners' limited English ability.

Converting the frequencies to percentages reveals a formal-informal ratio pattern (see Figure 4.5). This pattern, however, is different from the ratio discerned in the Additive MDMs, where the red bars (informal MDMs) were found to be taller than the blue bars (formal MDMs). In Figure 4.5, we observed taller blue bars (formal MDMs). In other words, learners at the three sub-corpora still employed more formal MDMs even when they wrote in informal style. Because of that, the ratio pattern observed here is identical to the ratio in the formal register (see Figure 4.4).

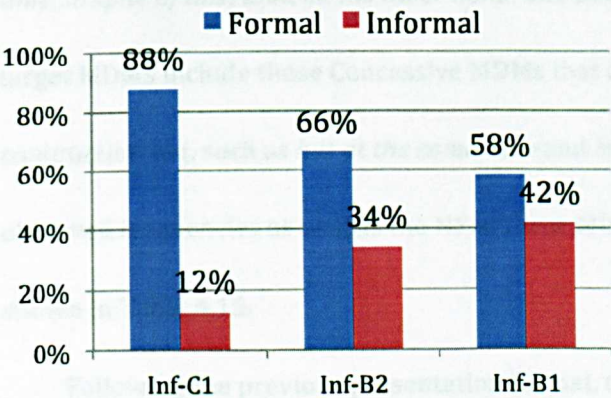


Figure 4.5 Percentages of formal and informal Causal MDMs, informal register

Despite the observed similarity between Figures 4.4 and 4.5, a close look at the differences between the two bar charts offers a new perspective. Comparing the two ratios, we see that B2 and B1 learners showed a decrease in the use of formal MDMs in the informal register, with decreases of 9 percent (formal register: 75% - informal register: 66%) and 25 percent (formal register: 88% - informal register: 58%)

respectively. The decreases in using formal MDMs, at the same time, led to the increases of 9 percent (34%-25%) and 25 percent (42%-17%) in learners' use of informal MDMs. In other words, even though learners still tended to use more formal MDMs in their informal writing, their proportional use of informal MDMs had increased when writing informally.

4.2.4 Concessive MDMs

4.2.4.1 Formal register

In the Concessive category, I searched seven MDMs: *despite this/that, having said that/this (that said, that being said), but then again (then again, but again), at the same time, in spite of this/that, on the other hand, and but/and still*. Other sub-types of the target MDMs include those Concessive MDMs that occur with *however* and the conjunction *but*, such as *but at the same time* and *but/however on the other hand*. The observed frequencies as well as the NF of these MDMs in the four sub-corpora are shown in Table 4.15.

Following the previous presentation format, the order of the MDMs in Table 4.15 is organised by formality. The first two MDMs are the most formal, while *on the other hand* and *at the same time* are the neutral MDMs, and the remaining three the most informal. C2 learners produced the most Concessive MDMs, whereas B1 learners used Concessive MDMs rather sparsely. All the learners from the four groups favoured the neutral MDM *on the other hand*, but they showed a difference in the least used MDMs. While C1 and B2 learners produced very few instances of the two informal MDMs, *but then again* and *having said that*, C2 learners employed a very small number of the

formal MDMs—*in spite of that* and *despite that*. B1 learners produced neither formal, nor informal MDMs but only used the neutral MDMs. B1 learners' zero frequency in the formal and informal Concessive MDMs might be explained by the following two accounts. First, B1 learners might have adopted the avoidance strategy. Alternatively, they might not be aware of the very formal and informal Concessive MDMs. In the face of the English proficiency at level B1 (IELTS score 4.0 to 5.0), it is probable that the latter explanation is the cause of such zero occurrences.

Table 4.15

RF and NF of Concessive MDMs in the four learner corpora, formal register

MDMs	Raw frequency				NF (per million words)			
	F-C2	F-C1	F-B2	F-B1	F-C2	F-C1	F-B2	F-B1
in spite of that	8	14	10	0	6.46	8.07	6.20	0.00
despite this/that	11	16	17	0	8.88	9.22	10.54	0.00
on the other hand	262	284	246	6	211.60	163.72	152.57	13.19
at the same time	47	22	13	1	37.96	12.68	8.06	2.20
but still	60	22	13	0	48.46	12.68	8.06	0.00
having said that	27	9	2	0	21.81	5.19	1.24	0.00
but then again	10	10	2	0	8.08	5.76	1.24	0.00
Sum					343.25	217.33	187.93	15.39

Despite the zero occurrences in B1, I have observed some frequency differences in other sub-corpora. Such differences were analysed via *LL* tests, shown in Table 4.16.

Table 4.16

LL test results of Concessive MDMs, formal register

MDMs	C2 vs. C1		C1 vs. B2		B2/B1	
	LL ratio	Sig. level	LL ratio	Sig. level	LL ratio	Sig. level
in spite of that	0.257	0.613, ns	0.409	0.522, ns	NA	
despite this/that	0.009	0.924, ns	0.148	0.701, ns	NA	
on the other hand	8.914	0.003 (C2)	0.654	0.418, ns	83.744	.000** (B2)
at the same time	16.649	.000** (C2)	1.729	0.189, ns	2.285	0.131, ns
but still	33.434	.000** (C2)	1.729	0.189, ns	NA	
but then again	0.566	0.452, ns	5.253	0.022, (C1)	NA	
having said that	16.506	.000** (C2)	4.321	0.038, (C1)	NA	

** p< .001 level

The first pair (C2 vs. C1) shows that C2 and C1 learners employed the formal MDMs in a similar manner, shown by the non-significant findings. Nevertheless, C2 learners showed a preference for the use of neutral and informal MDMs. We see a similar pattern in the pair 2 comparison in that although the two groups of learners used formal and neutral MDMs comparably, C1 learners showed a strong preference for informal MDMs. Putting pair 1 and pair 2 together, we see that the likelihood of producing informal MDMs increased as the level of English proficiency developed. I infer that the more advanced learners, like C2 learners in this case, took a more

aggressive approach in employing various MDMs in writing, hence the noticeable inclusion of the use of informal MDMs.

As the four groups of learners used a substantial number of neutral MDMs, I included the neutral MDMs in the formal-informal ratio analysis. Figure 4.6 shows the proportion of the formal, informal and neutral MDMs in the four sub-corpora.

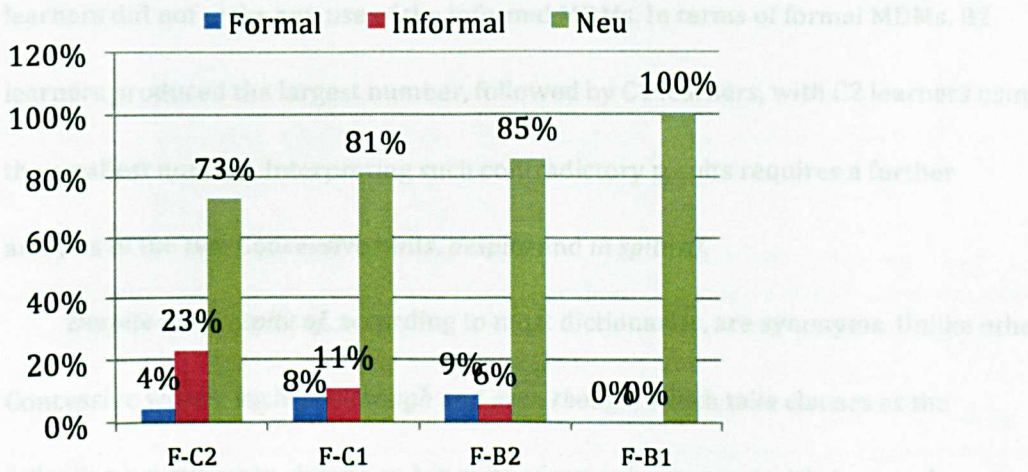


Figure 4.6 Percentages of formal, informal, and neutral Concessive MDMs, formal register

A first glance at the bar chart shows that the four groups of learners used a large portion of neutral MDMs, indicated by the long green bars. In addition, an upward trend was observed in the rate of neutral MDMs from the F-C2 corpus to the F-B1 corpus. This suggests that as learners' English level developed, the use of neutral MDMs decreased. Other than using the neutral MDMs, more advanced learners included the formal and informal MDMs in their writing.

A different pattern was observed from all the formal-informal ratios reported in the formal register thus far. In the sub-sections on Additive and Causal MDMs, C2 and C1 learners used more formal markers, whereas B2 and B1 learners produced more informal markers. In the case of Concessive MDMs, however, the trends of the blue (formal) and red (informal) bars developed in opposite directions—C2 learners used a proportionally greater number of informal MDMs, followed by C1 and B2 learners. B1 learners did not make any use of the informal MDMs. In terms of formal MDMs, B2 learners produced the largest number, followed by C1 learners, with C2 learners using the smallest number. Interpreting such contradictory results requires a further analysis of the two Concessive units, *despite* and *in spite of*.

Despite and *in spite of*, according to most dictionaries, are synonyms. Unlike other Concessive words, such as *although* and *even though*, which take clauses as the following constituents, *despite* and *in spite of* can only co-occur with nouns, phrases, or gerunds. In other words, although *despite that* and *in spite of that* are legitimate Concessive discourse markers, the occurrences of *despite* and *in spite of* are frequently found in the syntactic patterns with nouns, phrases, and gerunds. An additional search of *in spite of* and *despite* in the BNC-Acad corpus supported this line of argument. Of all the 104 retrieved instances of *in spite of*, the DM use (*in spite of this/that*) occurred with only 9 percent, whereas the non-DM use (in spite of Gerund/Noun/ Noun phrase) occurred with a relatively high proportion, 91 percent. There were 375 instances of *despite* retrieved from the BNC-Acad corpus and only 3 percent of them was used as DMs. The non-DM use of *despite* occurred with 97 percent.

Following the evidence found in the BNC-Acad corpus, I further searched *despite* and *in spite of* in the four learner corpora. This was to see whether or not learners exhibited a similar usage pattern. Some retrieved instances of *despite* and *in spite of* were cases of misuse. Even though learners' misuse is not discussed here, they were included in the results, shown in Table 4.17.

Table 4.17

Use of despite and in spite of as DMs and non-DMs in the four sub-corpora

	<i>despite</i>				<i>in spite of</i>			
	DM	non-DM	misuse	Sum	DM	non-DM	misuse	Sum
F-C2 (NF)	9	148	23	180	6	26	2	34
percentage	5%	82%	13%	100 %	19%	76%	5%	100 %
F-C1(NF)	9	152	42	203	8	42	9	59
percentage	4%	75%	21%	100 %	14%	72%	14%	100 %
F-B2(NF)	11	86	53	150	6	40	14	60
percentage	7%	57%	36%	100.00	10%	67%	23%	100 %
F-B1(NF)	0	2	2	4	0	9	2	11
percentage	0	50%	50%	100%	0	80%	20%	100 %

The non-DM use of *despite* and *in spite of* formed the majority of the instances retrieved. This is particularly noticeable when looking at the percentages. More than 50% of the instances found for *despite* and *in spite of* are used in the non-DM form, and this phenomenon was seen across the four sub-corpora. Learners' usage pattern with

regard to the DM and non-DM use of *despite* and *in spite of* was similar to that found in the BNC-Acad corpus. The finding also confirmed the line of reasoning discussed above. The conventional use of the two Concessive units, *despite* and *in spite of*, are their non-DM forms, and this can be attributed to the few instances of *despite that/this* and *in spite of that/this* retrieved in the four sub-corpora.

Going back to the earlier paragraph in which I mentioned the seemingly contradictory findings with regard to the ratios of informal and formal MDMs found in the Concessive category, I presume that the CLC learners' infrequent use of *despite* and *in spite of* in the DM form might account for the resulting increase in use of informal Concessive MDMs.

4.2.4.2 Informal register

The same group of Concessive MDMs was searched in the three sub-corpora in the informal register. The frequency information is shown in Table 4.18.

Similar to the finding in the formal register, learners across the three groups preferred the neutral MDM, *on the other hand*, as shown by the high frequency counts. The formal MDMs occurred with low frequency. The frequency of informal MDMs did not increase, although it is in the context of informal writing that these markers were examined. While C1 learners did not produce any instances of the very formal MDM, *in spite of that*, both B2 and B1 learners used some instances of this marker in their informal writing. The informal MDM *but still* was found to occur frequently in the formal register; in informal writing, however, it did not occur often.

Table 4.18

RF and NF of Concessive MDMs, informal register

MDMs	Raw frequency			NF (per million words)		
	Inf-C1	Inf-B2	Inf-B1	Inf-C1	Inf-B2	Inf-B1
in spite of that	0	3	6	0.00	5.28	5.46
despite this/that	4	7	0	39.57	12.31	0.00
on the other hand	12	63	36	118.70	110.79	32.74
at the same time	1	4	10	9.89	7.03	9.09
but still	1	4	2	9.89	7.03	1.82
but then again	0	1	0	0.00	1.76	0.00
having said that	0	0	0	0.00	0.00	0.00
Sum				178.04	144.20	49.11

To find out whether or not the frequency differences of the MDMs found in the three sub-corpora were significantly different, I applied *LL* tests. The results are presented in Table 4.19.

The non-significant differences in the pair one result (C1 vs. B2) showed that C1 learners and B2 learners used Concessive MDMs in a very similar way. A quite similar pattern is seen in the second pair comparison with the exception of the two groups' use of the marker *on the other hand*.

Table 4.19

LL test results of Concessive MDMs, informal register

MDMs	C1 vs. B2		B2 vs. B1	
	LL ratio	Sig. level	LL ratio	Sig. level
in spite of that	NA		0.002	0.962, ns
despite this/that	2.997	0.083, ns	NA	
on the other hand	0.047	0.828, ns	35.828	.000** (B2)
at the same time	0.087	0.768, ns	0.195	0.659, ns
but still	0.087	0.768, ns	2.639	0.104, ns
but then again	NA		NA	
having said that	NA		NA	

** p< .001 level

The formal-informal ratios of these MDMs are shown in Figure 4.7. Figure 4.7 showed a similar pattern of distribution to that of Figure 4.6, with some differences. Both Figures 4.6 and 4.7 include equally long green bars, which represent the neutral MDMs employed. The differences were in the increase and decrease of the blue and red bars. Regardless of the neutral MDMs, learners were found to produce more informal MDMs in formal writing. In the informal register, contrary to the expectation that learners might produce more informal MDMs, a higher proportion of formal Concessive MDMs was seen.

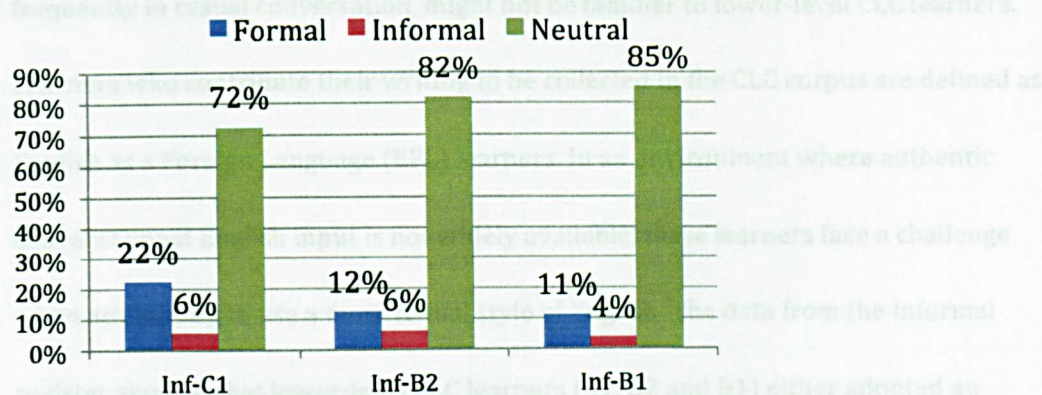


Figure 4.7 Percentages of formal, informal, and neutral Concessive MDMs, informal register

Piecing together the evidence from the informal register with the data observed in the formal writing, I propose that CLC learners showed problematic formality awareness in the case of Concessives. The frequently employed non-DM use of *despite* and *in spite of* explained the fact that learners produced fewer instances of formal Concessive MDMs in the formal register; yet, the same reason can be used to explain the fact that more instances of formal Concessive MDMs were found in the informal register. Because the use of *despite* and *in spite of* in non-DM forms involves a more complicated syntactic structure, such context-sensitive grammar structure might be beyond the reach of C1 and B learners. As a result, lower-level learners might opt for the DM use of *despite* and *in spite of*, which appear in sentences like frozen chunks, requiring no further variation in the forms of the following constituents.

Another interesting fact observed by comparing Figures 4.6 and 4.7 is the decrease of informal MDMs in informal writing. This can be ascribed to the fact that very informal Concessive MDMs (*having said that, but then again*), which occur very

frequently in casual conversation, might not be familiar to lower-level CLC learners. Learners who contribute their writing to be collected in the CLC corpus are defined as English as a Foreign Language (EFL) learners. In an environment where authentic conversational English input is not widely available, these learners face a challenge when attempting to use a more casual style of English. The data from the informal register showed that lower-level CLC learners (C1, B2 and B1) either adopted an avoidance strategy or they might simply not be aware of the existence of these informal Concessive MDMs.

4.2.5 Contrastive MDMs

4.2.5.1 Formal register

As mentioned in Chapter 3, the Contrastive is the only case in which all the MDMs were found to be either formal or neutral. The comparisons will thus be made mainly between the formal MDMs (*by/in contrast, on the contrary*), and the neutral MDM (*on the other hand*). The distributional patterns of observed frequencies and NF are shown in Table 4.20.

In the F-C2 and the F-C1 corpus, learners mostly used *on the contrary* and *on the other hand*. B2 and B1 learners preferred *on the other hand* to the other two MDMs. In other words, as is the case with Concessive MDMs, the neutral Contrastive MDM was the one favoured by most of the learners. This is a similar phenomenon to that observed in Chapter 3 where NS of English used the neutral Concessive and Contrastive MDMs more frequently than other MDMs. In terms of the total instances found, C2 learners used the most Contrastive MDMs and B1 learners used the fewest.

The least used MDM was the most formal: *by/in contrast*. The observed frequency differences were examined via *LL* tests. The results are shown in Table 4.21.

Table 4.20

RF and NF of Contrastive MDMs in the four learner corpora, formal register

MDMs	Raw frequency				NF (per million words)			
	F-C2	F-C1	F-B2	F-B1	F-C2	F-C1	F-B2	F-B1
by/in contrast	14	44	19	0	11.31	25.36	11.78	0.00
on the contrary	80	74	29	1	64.61	42.66	17.99	2.20
on the other hand	99	63	65	3	79.96	36.32	40.31	6.59
Sum					155.88	104.34	70.08	8.79

Table 4.21

LL test results of the Contrastive MDMs

MDMs	C2 vs. C1		C1 vs. B2		B2/B1	
	LL ratio	Sig. level	LL ratio	Sig. level	LL ratio	Sig. level
by/in contrast	7.821	0.005 (C1)	8.454	0.004 (C1)	NA	
on the contrary	6.614	0.010 (C2)	17.184	.000** (C1)	8.675	.000** (B2)
on the other hand	24.788	.000** (C2)	0.349	0.555, ns	16.804	.000** (B2)

** p< .001 level

In the pair one comparison, C2 learners used more instances of the neutral MDM *on the other hand*, and the formal MDM *on the contrary*, but C1 learners produced

more instances of the formal MDM *by/in contrast*. When contrasting C1 and B2 learners, I discovered that C1 learners used more instances of formal MDMs. At the same time, B2 learners used more formal and neutral MDMs than did B1 learners. Putting the three pairs together, I observed that a relationship seemed to exist between the number of formal Contrastive MDMs produced in the formal register and learners' English level—a tendency to produce formal Contrastive MDMs is associated with a higher level of English proficiency.

Since there are only formal and neutral MDMs in this category, I compared the ratios between the two types of MDM in the four sub-corpora, shown in Figure 4.8.

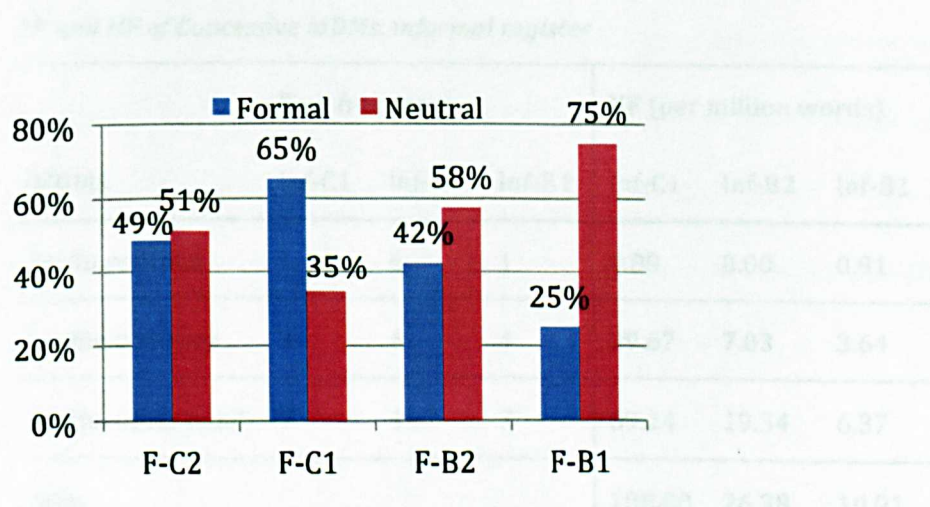


Figure 4.8 Percentages of formal and neutral Contrastive MDMs, formal register

Of all the Contrastive MDMs found in the F-C2 corpus, percentages of formal and neutral MDMs were evenly balanced. In the remaining three sub-corpora, however, we see once again the trend for a decrease in formal MDMs and an increase in neutral MDMs in less proficient learners. I calculated correlation coefficients using Pearson's method and found that there is a very high negative correlation between the formal MDMs and the neutral MDM ($r = -1.000$, $n = 4$, $p = 0.000$). This means that as these

learners' use of more formal MDMs decreased, their use of the neutral MDM increased.

This finding supports the previous claim that the level of English proficiency affects the number of formal MDMs produced by English learners.

4.2.5.2 Informal register

The same three Contrastive MDMs were searched in the Inf-C1, the Inf-B2, and the Inf-B1 sub-corpus. A summary of the frequencies is presented in Table 4.22.

Table 4.22

RF and NF of Concessive MDMs, informal register

MDMs	Raw frequency			NF (per million words)		
	Inf-C1	Inf-B2	Inf-B1	Inf-C1	Inf-B2	Inf-B1
by/in contrast	1	0	1	9.89	0.00	0.91
on the contrary	3	4	4	29.67	7.03	3.64
on the other hand	7	11	7	69.24	19.34	6.37
Sum				108.80	26.38	10.91

The formality-neutral MDM *on the other hand* appeared to be produced the most frequently, as is the case in the formal register. C1 learners used more Contrastive MDMs than did the other two B-level learner groups.

LL tests were again carried out to see whether the frequency differences reached statistical significance. The results are presented in Table 4.23.

Table 4.23 *Results in Figures 4.8 and 4.9 showed some interesting patterns.*

LL test results of Contrastive MDMs, informal register

MDMs	C1 vs. B2		B2 vs. B1	
	LL ratio	Sig. level	LL ratio	Sig. level
by/in contrast	NA		NA	
on the contrary	3.093	0.079, ns	0.854	0.355, ns
on the other hand	6.014	0.014 (C1)	5.455	0.020 (B2)

** p< .001 level

The paired comparisons showed that the difference in using Contrastive MDMs lay exclusively in the three groups' use of *on the other hand*. While C1 learners made use of *on the other hand* significantly more than did B2 learners, B2 learners also used *on the other hand* more frequently than did B1 learners. The formal-neutral ratios in each sub-corpus are shown in Figure 4.9.

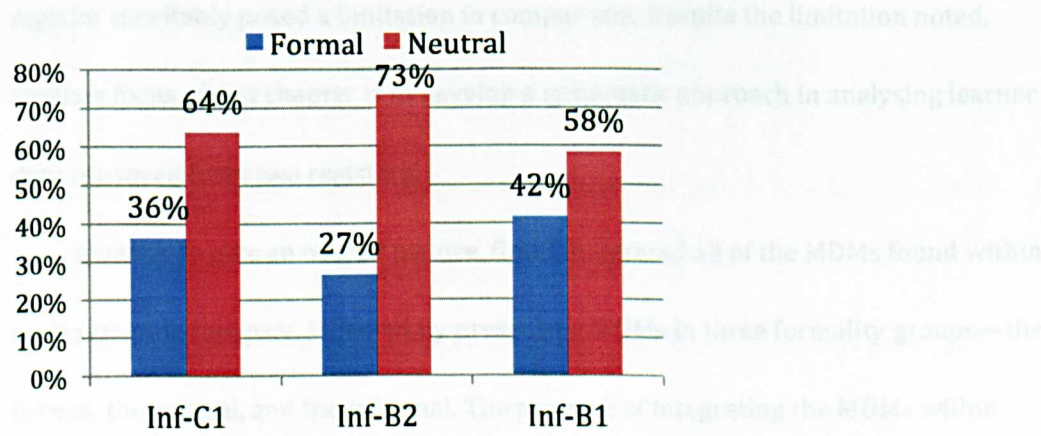


Figure 4.9 Percentages of formal and neutral Contrastive MDMs, informal register

The ratios in Figures 4.8 and 4.9 showed some interesting patterns.

Proportionally, both C1 and B2 learners decreased their use of formal MDMs in the informal register, with C1 learners' rate dropping from 65% in the formal register to 36% in the informal register, and B2 learners' rate dropping from 42% to 27%. The decreases in turn were reflected by an increase in the frequency of the neutral MDM in the informal register. Although B1 learners seemed to show an increase in the proportion of formal MDMs in the informal register, they only made use of a very limited number of Contrastive MDMs.

4.2.6 Putting corpus data together

I have presented the corpus data retrieved from the formal and informal written texts of English learners. The comparisons between the formal and the informal register were undertaken primarily within each semantic category. Note that the number of MDMs retrieved from the informal register was smaller than the number found in the formal register. The underrepresentation of MDMs in the informal register inevitably posed a limitation in comparison. Despite the limitation noted, another focus of this chapter is to develop a systematic approach in analysing learner data retrieved from two registers.

In order to give an overall picture, first I integrated all of the MDMs found within each semantic category, followed by presenting MDMs in three formality groups—the formal, the neutral, and the informal. The purpose of integrating the MDMs within each semantic category is to provide a global view of the retrieved data, and to see to what extent English learners' level of proficiency affects the quantity of MDMs

produced. The breakdown of the MDMs in accordance with the formality weight, on the other hand, offers a window into the relationship between the English proficiency level and the quality (with regard to the formality) of MDMs produced.

The integration information is shown in Table 4.24 and Table 4.25, with the former displaying the data from the formal register, and the latter the figures from the informal register.

Table 4.24

Summary of the total MDMs in each semantic category, formal register

MDMs	F-C2 (NF)	F-C1 (NF)	F-B2 (NF)	F-B1 (NF)	Sum 1
Additive	573.43	811.1	782.72	474.76	2642.01
Appositive	133.26	21.91	24.81	10.99	190.97
Causal	415.13	262.87	168.08	65.94	912.02
Concessive	343.25	217.33	187.93	15.39	763.90
Contrastive	155.88	104.34	70.08	8.79	339.09
Sum 2	1620.95	1417.55	1233.62	575.87	

Sums 1 and 3 show the quantity of MDMs that learners produced within each semantic category of MDMs. Ranking the five categories of MDMs in order of frequency, I observed that the Additive category was used the most frequently, whereas the Appositive was used the least. The Causal and Concessive categories formed the median group, and the Contrastive was ranked the fourth most frequently used in both registers. The ranking order in the formal register is similar to that of the

informal register, suggesting that CLC learners might find the Additive concept easier than other types of MDM. At the same time, CLC learners might find using the Appositive MDMs the most difficult, or they simply avoid using Appositive MDMs.

Table 4.25

Summary of the total MDMs in each semantic category, informal register

MDMs	Inf-C1 (NF)	Inf-B2 (NF)	Inf-B1 (NF)	Sum 3
Additive	722.06	650.65	224.64	1597.35
Appositive	19.78	12.31	3.64	35.73
Causal	79.13	103.75	39.11	221.99
Concessive	178.04	144.2	49.11	371.35
Contrastive	108.8	26.38	10.91	146.09
Sum 4	1107.81	937.29	327.41	

Although the two tables include the frequency information from two registers, Sum 2 in Table 24 bears a striking resemblance to Sum 4 in Table 25. That is, higher-level learners tended to produce more MDMs than did lower-level learners. As this phenomenon applies to both the formal and the informal register, I conclude that English levels do affect the quantity of MDMs that CLC learners produce.

A further question is whether English level also plays a role in the formality types of MDMs produced. To answer this, I have further subcategorized the NF of MDMs by formality. The results are shown in Figures 4.10 and 4.11, with the former showing the breakdown in the formal register, and the latter the summary in the informal register.

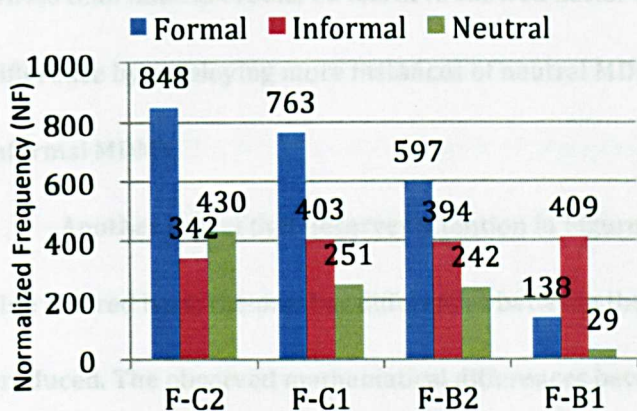


Figure 4.10 Breakdown of MDMs by formality category, formal register

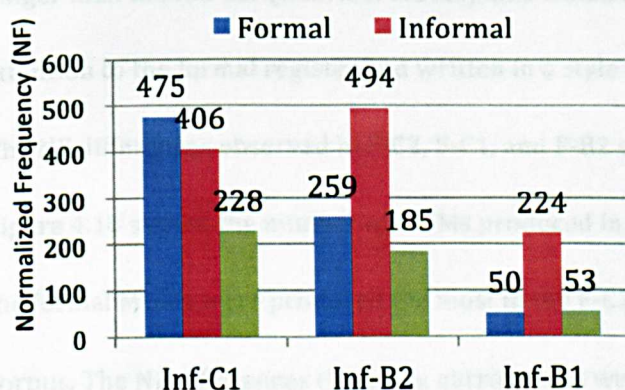


Figure 4.11 Breakdown of MDMs by formality category, informal register

In Figure 4.10, we observed a gradual decrease in both the blue (the formal MDMs) and the green (the neutral MDMs) bars from the F-C2 corpus to the F-B1 corpus. The declining trend in NF suggests that more advanced learners tended to use more formal and neutral MDMs. The number of informal MDMs, on the other hand, showed a different trend. An upward shift in NF from the F-C2 corpus to the F-B1 corpus was seen. In other words, the use of informal MDMs increased with lower level

learners of English. In addition, while C1, B2 and B1 learners produced more informal MDMs than neutral MDMs, C2 learners showed better noticing of the formality difference by employing more instances of neutral MDMs and fewer examples of informal MDMs.

Another aspect that deserves attention in Figure 4.10 is the differences between blue and red bars: the number differences between the formal and informal MDMs produced. The observed mathematical differences between the two bars indicate how much attention learners have paid to the register difference. There are two scenarios when examining these differences. To start with, when blue bar (formal MDMs) is longer than the red bar (informal MDMs), this would suggest that learners have paid attention to the formal register and written in a style that conforms to that register. The NF differences observed in F-C2, F-C1, and F-B2 are examples of this case. As Figure 4.10 shows, the number of MDMs produced in the formal register, we see that the formal MDMs were produced the most in the F-C2 corpus, and the least in the F-C1 corpus. The NF differences therefore narrowed as we moved along from the F-C2 corpus (NF difference of 506 between formal and informal MDMs), to the F-C1 corpus (NF difference of 360), to the F-B2 corpus (NF difference of 203). From these facts we can infer that a relationship exists between the differences observed and the amount of pragmatic awareness to registers that learners have developed. C2 learners in this case have developed, comparatively speaking, a greater awareness, as they tended to produce far more formal MDMs than informal MDMs in the formal register.

The second scenario is when the difference is a result of the subtraction of a shorter blue bar from a longer red bar, indicating that more informal MDMs are

employed. Because it is in the formal register where the subtraction occurs, the gap here indicates the extent to which learners are unaware of the formality. The blue-red bar gap in the case of the F-B1 corpus (NF difference of 271) belongs to this category. B1 learners, when compared with the other three groups, showed a lack of register sensitivity. Considering the NF differences in the four sub-corpora, we can infer that the pragmatic awareness in using MDMs in accordance with the formal register develops as learners' English level develops. C2 learners evidently demonstrated a keen awareness in the formal register.

The bar chart in the informal register (Figure 4.11) shows a different outlook. Although the blue bars (the formal MDMs) still showed a declining trend, the red bars (informal MDMs) displayed a different trend. B2 learners produced far fewer formal MDMs and at the same time, a larger number of informal MDMs, which resulted in a wider formal-informal gap, NF: 235 (494-259). C1 learners still produced more formal MDMs in the informal register, yet the formal-informal gap had decreased to NF: 69 (475-406). B1 learners still employed a large number of informal MDMs (the red bar).

Piecing together the evidence shown in Figures 4.10 and 4.11, it seems almost certain that, when looking at all the MDMs by formality and register, C2 learners have shown a sound grasp of pragmatic awareness in using the MDMs in accordance with the register difference. The breakdowns of the MDMs in Inf-C1 (Figure 4.11) have nevertheless clouded the view of C1 learners' seemingly reasonable awareness in the formal register. Both B2 and B1 learners have used more informal MDMs in the informal register, which on the face of it might suggest their awareness in the informal register. Such an inference, however, becomes invalid if we include the information

shown in Figure 4.10. B1 learners clearly showed a lack of register awareness in both the formal and the informal register.

Note also that the four groups of learners preferred the neutral Contrastive and Concessive MDM *on the other hand*. Such usage preference was evident in both the NS and the L2 data. It seems that when language users make use of MDMs, two factors are affecting their choices: the formality weight and the frequency effect. Depending on the register, or on how formal/informal a context is, the formality weight affects language users' choice of MDMs. At the same time, the frequency effect also plays a role in such a decision; that is, the more frequently an MDM is used, the more likely it is that this MDM will appear in language users' output regardless of the context.

Examining both NS and L2 writing in the formal register, I have discovered that the most formal MDMs were not necessarily produced most frequently. As explained, the frequency effect also influences language users' choice. Each of the two factors, to some extent, exemplifies the dynamic nature of language use. The formality weight of the Contrastive and Concessive MDM *on the other hand* suggests that it is a neutral-toward-formal MDM. But as it is a frequently occurring phrase, both NS and learners are more likely to use this phrase in both the formal and the informal context.

4.2.7 Summary

I have presented the CLC data in accordance with the semantic categories in the formal and informal registers. I have examined CLC learner data both quantitatively and qualitatively. In what follows, I summarize the key findings from this section.

1. English level plays an influential role in not only the quantities but also

the varieties (formality and semantic category) of MDMs that learners produce.

2. Converting the total NF to a percentage in each sub-corpus in the formal register, C2 learners showed a high ratio of formal MDMs to informal MDMs. B2 learners, on the other hand, showed a high ratio of informal to formal MDMs.
3. In the formal register, more advanced learners demonstrated their pragmatic competence with regard to the discourse and formality level by employing more formal MDMs.
4. In the informal register, learners tended to produce fewer instances of MDMs in general, but more informal MDMs were found in learners' informal writing.

The findings thus far provide answers to the third and fourth research questions that I raised in Chapter 1:

RQ3. Does learners' insensitivity to formal register difference apply to spoken/less formal discourse?

RQ4. Does the English proficiency affect L2 writers' use of MDMs in terms of quantity, varieties, and types?

With regard to research question 3, the findings suggest that, contrary to what is reported in the literature, L2 learners do not have problems in using formal MDMs in

the formal register. Although the CLC learners were found to be less sensitive in using appropriate MDMs in the informal register, this phenomenon did not apply to all of the five semantic types of MDM. With regard to research question 4, the data has shown that English proficiency does play a role in deciding how many, what semantic type, and which formality types of MDMs L2 learners use. Advanced L2 learners use more MDMs and yet their choice of MDMs conforms to the formality level of the context.

4.3 Comparing the learner data with NS data

In Section 4.2, I created seven sub-corpora on the basis of learners' English level. The purpose was to make a cross-section comparison, to see how English level correlated with the quantity and quality of MDMs produced by learners. In this section, the purpose is to directly compare learner data with the four NS reference corpora: the BNC-Acad corpus, which represents the written, formal register; the BASE corpus, which characterizes the spoken, formal register; the Enronsent2 e-mail corpus, which shows the written, informal register; and the CANCODE corpus, which stands for the spoken, informal register (see Chapter 3).

Instead of comparing the seven learner sub-corpora with the four NS corpora directly, the seven sub-corpora were merged in accordance with the CEFR levels. For example, the F-C2 and the F-C1 corpus were combined together to form the F-C corpus, which is the database that archived C-level learners' writing in the formal register. The combination is to avoid repetition of the data reported in Section 4.2, and further, to draw a generalizable inference, relative to the use of MDMs. Table 4.26 shows the metadata of the merged learner corpora and the four NS corpora.

Table 4.26

Summary of the corpora to be compared

Corpus	Nature	Size (words)
BNC-Acad	Written Eng. (academic, formal)	2,074,185
BASE	Spoken Eng. (academic, formal)	1,186,290
Enronsent2	Written Eng. (personal, informal)	9,826,382
CANCODE	Spoken Eng. (conversational, informal)	4,859,392
F-C	C learner writing, formal register	2,972,856
F-B	B learner writing, formal register	2,067,300
Inf-C	C learner writing, informal register	101,099
Inf-B	B learner writing, informal register	1,668,184

The discussion of the comparisons in this section is less concerned with the overuse/underuse phenomenon. I will focus on the formal-informal ratios between the learner data and the NS data. The presentation of the comparisons will follow the format in Section 4.2: presenting the MDMs in accordance with the semantic categories. Instead of showing the NF of each MDM within each semantic category, I clustered together MDMs that suggested the same level of formality. For example, *in addition* and *add to this* were clustered to form the formal category, whereas *what's more*, *on top of that*, *not only that*, and *and also*, were gathered together to form the informal category. For a clearer representation, the cumulative proportion of formal,

neutral, and informal MDMs will be presented as a stacked column chart totalling 100% in each semantic category.

4.3.1 Additive MDMs

The seven Additive MDMs are subcategorized into three categories: formal (*in addition, add to this*), neutral (*besides this*), and informal (*and also, what's more, on top of that, and not only that*). The NF in the four sub-corpora and the four NS corpora are shown in Table 4.27.

Table 4.27

NF of Additive MDMs in the NS and learner corpora, F=formal, Neu=Neutral, Inf=informal

	F. MDMs	Neu. MDMs	Inf. MDMs	
Corpus	(NF of <i>in addition</i>)		(NF of <i>and also</i>)	Sum
BNC-Acad	78.10 (76.17)	0.48	76.17 (71.84)	154.76
BASE	7.59	0.84	155.95 (146.68)	164.38
Enronsent2	93.93 (93.73)	0.31	42.34 (39.89)	136.57
CANCODE	1.65	0.21	86.43 (77.79)	88.28
F-C	383.47 (373.77)	26.91	301.73 (246.23)	712.11
F-B	337.15 (334.25)	23.70	354.09 (330.87)	714.94
Inf-C	316.52 (316.52)	19.78	385.76 (197.83)	722.06
Inf-B	64.74 (63.54)	16.19	288.94 (245.18)	369.86

CLC learners' overuse of Additive MDMs seems at first glance to be obvious.

Although learners' overuse/underuse is not the focus of the comparison, it should be noted that the same overuse phenomenon in Additive MDMs is seen in other semantic categories, with the exception of the Appositive MDMs. Such overwhelming overuse of MDMs is due largely to the fact that, unlike NS, who possess and exercise linguistic skills intuitively, L2 learners have a limited repertoire of linguistic resources. Various MDMs are therefore employed by L2 learners to compensate for their not yet mastered linguistic forms.

One similarity was observed between the two written NS corpora and the learner corpora. In the BNC-Acad and the Enronsent2 corpus, NS writers seemed to prefer the formal Additive MDM *in addition* and the informal MDM *and also*. Such preference was also seen in the four learner groups. The NF of *in addition* and *and also* was indicated by the round brackets in Table 4.27.

In order to see the proportions of the total MDMs produced in the eight corpora, I converted the NF in Table 4.27 to percentages. Despite learners' overuse of MDMs in general, the conversion provides a chance to see whether learners' choices of the formal, neutral, and informal MDMs bear resemblances to those of NS. Figure 4.12 shows the cumulative proportion of formal, neutral, and informal MDMs.

Despite the fact that the two learner groups in the F-C and F-B corpora have produced four times more Additive MDMs than the NS in the BNC-Acad corpus, both C- and B-level learners displayed similar formal-informal ratios to those in the BNC-Acad corpus. The two learner corpora in the informal register (the Inf-C and the Inf-B corpus) showed a different pattern. Forty-four percent of the Additive MDMs retrieved

from the Inf-C corpus were formal MDMs. B learners, on the other hand, produced a lot more informal MDMs in the Inf-B corpus. The pattern observed in the Inf-B corpus approximated more closely to that observed in the CANCODE corpus. Another observed phenomenon was CLC learners' use of the neutral Additive MDM *besides that*. The four NS corpora, on the other hand, produced rather few instances of the neutral MDM.

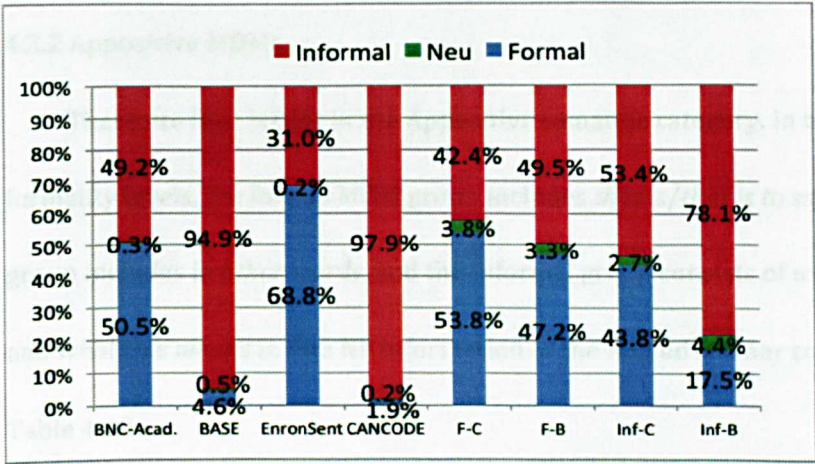


Figure 4.12 Formal-informal ratios of Additive MDMs

It should be noted that even though the CANCODE corpus and the two learner corpora (the Inf-C and the Inf-B corpus) archive writing in the informal register, the language modes are different, with the former being in the spoken mode and the latter two in the written form. The difference in modes suggests that a direct comparison between the two presents the risk of a faulty inference. That is, while it is possible to say that B-level learners showed a greater awareness of register difference by using more informal MDMs in the Inf-B corpus, it might also be due to the fact that the written mode has attributed to the more formal MDMs found in the Inf-C corpus.

A tentative conclusion drawn from the corpus evidence is that both the B-level and C-level learners have displayed a firm grasp of Additive MDMs in the formal register. In the informal register, however, more advanced learners (C-level learners) still produced a large portion of formal MDMs. C-level learners' preference for formal MDMs in the informal register is also seen in other semantic categories, which will be elaborated further in the following sections.

4.3.2 Appositive MDMs

There are four MDMs in the Appositive semantic category. In terms of the formality levels, the formal MDM group includes *that is/that is to say*, the neutral group includes *in other words*, and the informal group consists of *what I'm saying is* and *what this means is*. The NF information of the NS and learner corpora are shown in Table 4.28.

Of all the five semantic categories, the only one that showed learners' underuse is Appositives, indicated by the sum column in Table 4.28. NS in both the BNC-Acad and the BASE corpus produced many instances of formal Appositive MDMs. In the informal context, NS in the Enronsent2 and the CANCODE corpus used more of the neutral Appositive MDM *in other words*.

Contrary to NS's usage pattern in this regard, learners in the four corpora used a lot fewer instances of Appositive MDMs, regardless of whether the register was formal or informal. While the NS in the BNC-Acad corpus employed a large portion of formal Appositive MDMs, they also made frequent use of the neutral Appositive MDM *in other words*. In this regard, only advanced learners in the F-C corpus showed a similar

pattern. While NS used more neutral appositive MDMs in the informal register, learners in the Inf-C and Inf-B corpus did not show such a tendency.

Table 4.28

NF of Appositive MDMs in the NS and learner corpora, F=formal, Neu=Neutral, Inf=informal

Corpus	F. MDMs	Neu. MDMs	Inf. MDMs	Sum
BNC-Acad	139.33	70.39	0.48	264.20
BASE	79.24	21.92	39.62	140.78
Enronsent2	9.36	13.33	1.53	24.22
CANCODE	2.40	31.80	3.20	37.40
F-C	30.61	37.34	0.34	68.29
F-B	12.09	9.67	0.00	21.77
Inf-C	9.89	9.89	0.00	19.78
Inf-B	3.00	3.00	0.60	6.60

The formal-neutral-informal ratios in the eight corpora are shown in Figure 4.13. The formal-informal ratios in the four NS corpora showed that while the proportion of formal Appositive MDMs (the blue bars) decreased as the register changed from formal, written (BNC-Acad) to informal, spoken (CANCODE), the percentages of the neutral and informal Appositive MDMs increased. In spite of the very few instances of Appositive MDMs found in the four learner corpora, the four groups of learners produced mainly formal and neutral Appositive MDMs, suggesting that neither C-level

nor B-level learners noticed the formality difference between the formal and the informal register.

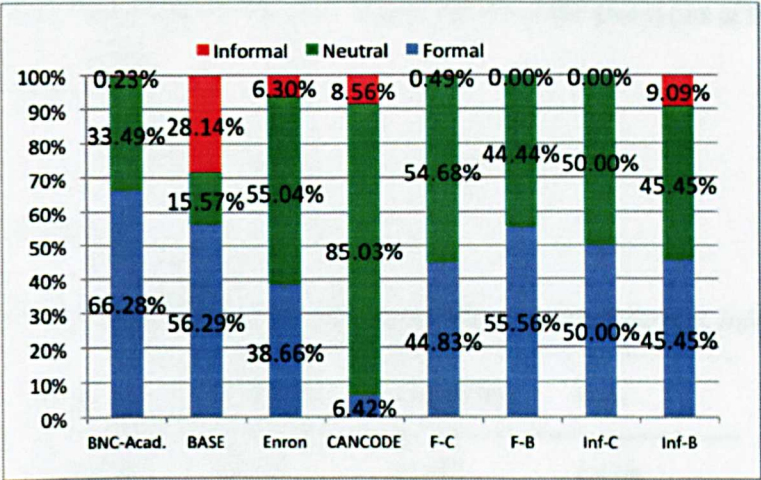


Figure 4.13 Formal-neutral-informal ratios of Appositive MDMs

Comparing the ratios in Figure 4.13 and the NF information in Table 4.28, I discovered that the use of Appositive MDMs was more problematic for learners. The unchanged formal-neutral-informal ratios in both the formal and the informal register suggest that learners in the CLC corpus did not pay particular attention to the formality levels when using Appositive MDMs. The very low frequencies found in the four learner corpora also suggest that learners might adopt an avoidance strategy with regard to Appositive MDMs. The evidence gathered from the NF of Appositive MDMs and the ratio information suggests that CLC learners seem to face a rather challenging task in employing Appositive MDMs.

4.3.3 Causal MDMs

Four MDMs are included in this category and they are further categorized into two groups: formal (*as a result, as a consequence, and for that reason*), and informal (*because of that*). Table 4.29 shows the NF of the two types of MDM in the eight corpora.

Table 4.29

NF of Causal MDMs in the NS and learner corpora, F=formal, Inf=informal

Corpus	F. MDMs	Inf. MDMs	Sum
BNC-Acad	92.57	7.23	99.80
BASE	26.13	10.12	36.25
Enronsent2	30.94	2.44	33.38
CANCODE	6.17	5.35	11.52
F-C	296.68	29.60	326.29
F-B	110.29	35.31	145.60
Inf-C	69.24	9.89	79.13
Inf-B	38.37	22.78	61.14

The total number of Causal MDMs produced in the four NS corpora declined from NF: 98.00 in the very formal written register (BNC-Acad) to NF: 11.52 in the very informal spoken register (CANCODE). Learners also showed a similar yet slightly different pattern in this regard. The similarity lay in the fact that another declining trend was observed in the sum of the Causal MDMs—from the high level learners in

the formal register (F-C) to the lower level learners in the informal register (Inf-B).

While the F-C and F-B corpora showed an overuse of Causal MDMs, the two learner corpora in the informal register did not show such a tendency. In the informal register, both C-level and B-level learners used more formal MDMs; yet as has been observed in the section on Additive MDMs, learners in the Inf-C corpus produced more formal Causal MDMs than did B-level learners.

The formal-informal ratios of the Causal MDMs produced in the eight corpora are shown in Figure 4.14.

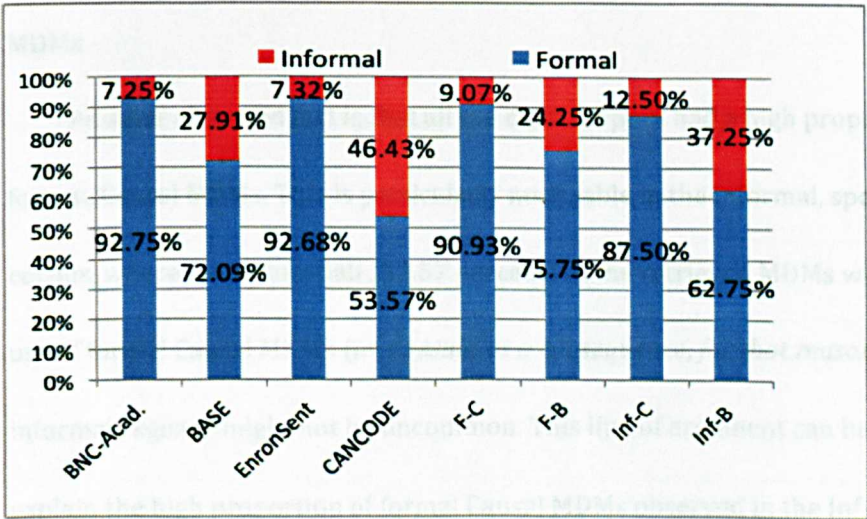


Figure 4.14 Formal-informal ratios of Causal MDMs

The ratios in Figure 4.14 showed some interesting facts. The ratio of the F-C learner corpus bore a striking resemblance to that of the BNC-Acad corpus as well as that of the Enronsent2 corpus. The ratios observed in the previous two semantic categories did not show any pattern of likeness between the F-C and the Enronsent2 corpus. The ratios observed between the F-C and the BNC-Acad corpus, on the other

hand, bore a general resemblance. The similarity between the F-C and the Enronsent2 e-mail corpus might therefore be a chance result. One inference drawn from the similarity observed between the BNC-Acad and the F-C corpus is that C-level learners show a greater level of noticing of the register difference in the formal writing discourse. Another similarity was found between the ratios in the F-B (75.75% to 24.25%) and the BASE corpus (72.09% to 27.91%), indicating that B-level learners' use of Causal MDMs in the formal register closely resembled that of NS in the formal spoken register. The two comparisons, F-C vs. BNC-Acad and F-B vs. Enronsent2, lead to the belief that the level of English proficiency has an influence upon learners' use of MDMs.

Another observed fact is that all the eight corpora had a high proportion of formal, Causal MDMs. This is particularly noticeable in the informal, spoken CANCODE corpus, where more than half (53.57 percent) of the retrieved MDMs were formal. The use of formal Causal MDMs (*as a result, as a consequence, for that reason*) in the informal register might not be uncommon. This line of argument can be used to explain the high proportion of formal Causal MDMs observed in the Inf-C (82.5%) and the Inf-B (62.75%) corpus. As is the case with the ratios in the previous two semantic categories, the ratio observed in the Inf-C corpus showed that C-level learners mainly made use of formal MDMs. Their use of informal Causal MDMs in the informal register was rarely observed. The phenomenon of C-level learners' preference for formal MDMs in the informal register, which repeatedly occurred in the Additive, Appositive, and Causal semantic categories, might suggest that advanced learners are more accustomed to writing in a formal style. Their unfamiliarity with writing in a less

formal style might result in the higher frequency of formal MDMs. B-level learners, on the other hand, did not show such a tendency.

4.3.4 Concessive MDMs

The seven Concessive MDMS are further divided into three groups—the formal group (*in spite of this/that, despite this/that*), the neutral group (*on the other hand, at the same time*), and the informal group (*but still, but then again, and having said that*).

The NFs of the three groups in the eight corpora are shown in Table 4.30.

Table 4.30

NF of Concessive MDMs in the NS and learner corpora, F=formal, Neu=Neutral,

Inf=informal

Corpus	F. MDMs	Neu. MDMs	Inf. MDMs	Sum
BNC-Acad	9.16	69.91	25.07	104.14
BASE	2.53	32.03	128.97	163.54
Enronsent2	0.61	6.31	21.07	27.99
CANCODE	1.03	19.76	222.66	243.45
F-C	16.48	206.87	46.42	269.77
F-B	13.06	128.67	8.22	149.95
Inf-C	39.57	128.59	9.89	178.04
Inf-B	9.59	67.74	4.20	81.53

While most of the learners showed an overuse of Concessive MDMs, learners in the Inf-B corpus did not. In terms of the formality group of the MDMs, the NS writers in the BNC-Acad corpus used more neutral Concessive MDMs, whereas NS in the other three NS corpora employed more informal Concessive MDMs. The preference for using neutral Concessive MDMs was also seen in the four learner corpora. Informal, Concessive MDMs, on the other hand, occurred rather infrequently in the learner corpora, except in the F-C corpus.

In order to discover the proportional differences in using Concessive MDMs at various formality levels, I converted the NF of the three groups to percentages. The formal-neutral-informal ratios of the Concessive MDMs in the eight corpora are shown in Figure 4.15.

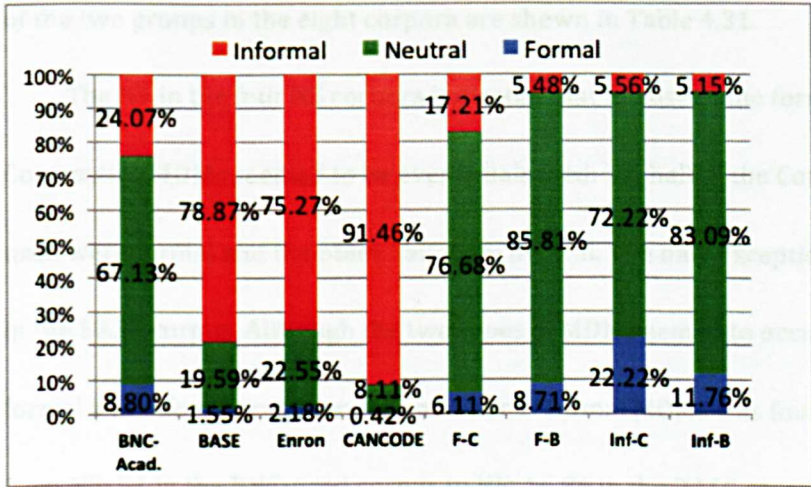


Figure 4.15 Formal-neutral-informal ratios of Concessive MDMs

The fact that the formal Concessive MDMs occurred at a frequency of approximately 9 percent in the BNC-Acad corpus might seem unusual at first glance.

Nevertheless, as explained in Section 4.2.4, the non-DM use of *in spite of* and *despite* occur far more frequently than the DM use of these two Concessive units. The two learner corpora in the formal register seemed to follow this usage pattern. Learners in the informal register, however, were found to use a higher ratio of formal MDMs, with a relatively limited use of informal MDMs. This finding suggests that both C- and B-level learners have greater difficulties in using Concessive MDMs in the informal register.

4.3.5 Contrastive MDMs

Unlike the previous semantic categories, the Contrastive category has only formal and neutral contrastive MDMs. The formal group consists of two MDMs, *by/in contrast*, and *on the contrary*, whereas the neutral group includes only *on the other hand*. The NF of the two groups in the eight corpora are shown in Table 4.31.

The NF in the four NS corpora indicated that the use of the formal and the neutral Contrastive MDMs seemed to be evenly balanced: i.e., half of the Contrastive MDMs used were formal and the other half were neutral. The only exception is the NF found in the BASE corpus. Although the two types of MDM seemed to occur evenly in the formal and informal registers, the number of formal MDMs was found to decrease from NF: 54 in the BNC-Acad corpus to NF: 16.86 in the BASE corpus, NF: 3.26 in the Enronsent2 corpus, and to NF: 2.06 in the CANCODE corpus. The decreasing trend was also observed in the NF of neutral MDMs. The downward trend in the number of Contrastive MDMs produced is, in itself, an indication of the formality level, as the number of Contrastive MDMs was greatest in the formal, written register (the

BNC-Acad corpus) and smallest in the informal, spoken register (the CANCODE corpus). That is, the total number of Contrastive MDMs used in the four NS corpora serves as an indication of register.

Table 4.31

NF of Contrastive MDMs in the NS and learner corpora, F=formal, Neu=Neutral

Corpus	F. MDMs	Neu. MDMs	Sum
BNC-Acad	54.00	54.00	108.00
BASE	16.86	28.66	45.52
Enronsent2	3.26	3.26	6.52
CANCODE	2.06	3.09	5.14
F-C	71.31	54.49	125.80
F-B	23.70	32.89	56.60
Inf-C	39.57	69.24	108.80
Inf-B	5.40	10.79	16.19

The NF in the four learner corpora showed a different pattern. C-level learners produced more Contrastive MDMs than did B-level learners in both the formal and the informal register. In addition, the F-C corpus was the only database where more formal MDMs were found. More neutral MDMs were retrieved in the other three learner corpora. In terms of the total number, B-level learners produced the fewest instances of Contrastive MDMs in the informal register (Inf-B).

I converted the NF of the eight corpora to percentages and this information is shown as a bar chart in Figure 4.16.

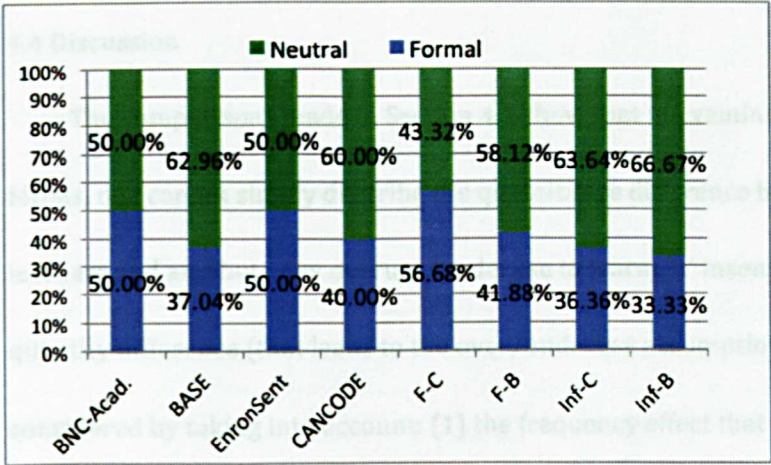


Figure 4.16 Formal-neutral ratios of Contrastive MDMs

The formal-neutral ratios in the figure offered a clear picture of the formality segmentation of the Contrastive MDMs. While both the formal and neutral MDMs were employed in the four NS corpora, the neutral MDMs occurred with a higher proportion in the spoken register, shown by the longer green bars in the BASE and CANCODE corpora. The proportion of the formal Contrastive MDMs used in the four learner corpora showed a declining trend from 57% in the F-C corpus to 42% in the F-B corpus, 37% in the Inf-C corpus, and 33% in the Inf-B corpus. At the same time, the percentage of the neutral Contrastive MDM *on the other hand* increased. This observed change in the formal-neutral ratios might be seen as an indication that learners' English level plays a role in not only the quantity but also the types of Contrastive MDMs that learners use. In addition, similar formal-neutral ratios were observed in the BASE and the Inf-C corpus—37%-63%—suggesting that C-level learners' use of

Contrastive MDMs in the informal register approximated to NS' use of these MDMs in the formal, spoken register.

4.4 Discussion

The comparisons made in Section 4.3 show that in examining learners' use of MDMs, one cannot simply describe the quantitative difference between NS and learners and attribute any overuse/underuse to learners' insensitivity to register. The quantity difference (that leads to the over/underuse assumption) should be considered by taking into account: (1) the frequency effect that different semantic types (of MDMs) might have on the number of MDMs L2 learners produce, and (2) the formal-informal ratios within the measured quantity. With a refined analysis of the quantity, we see that the frequencies of various semantic types of MDM suggest the ease or difficulty that learners face when using specific MDMs. In what follows, I will present the discussion by focusing on two aspects: the frequency effect that lies within each semantic type of MDMs, and the formality makeup of the MDMs.

4.4.1 The frequencies of MDMs in the five semantic categories

Earlier I presented the frequency information for the five semantic categories and carried out the analysis and comparisons between the NS and the learner data in each semantic type of MDM. In this section, I provide a global view by presenting the frequencies of the five semantic types of MDM in the same table. Any NS-learner differences observed from the summaries of the NF distribution would provide insights into how L2 learners acquire different semantic types of MDM.

The total NF counts of NS data are shown in Table 4.32 and the learner data are summarized in Table 4.33.

Table 4.32

A summary of all the MDMs in each semantic category: NS data

	Additive	Appositive	Causal	Concessive	Contrastive	Sum 1
Corpus	(NF)	(NF)	(NF)	(NF)	(NF)	(NF)
BNC-Acad	154.76	210.20	99.80	104.14	107.99	676.89
BASE	164.38	140.78	36.25	163.54	45.52	550.47
Enronsent2	136.57	24.22	33.38	27.99	6.51	228.67
CANCODE	88.28	37.40	11.52	243.45	5.14	385.79
Sum 3 (NF)	543.99	412.60	180.95	539.12	165.16	

Table 4.33

A summary of all the MDMs in each semantic category: learner data

	Additive	Appositive	Causal	Concessive	Contrastive	Sum 2
Corpus	(NF)	(NF)	(NF)	(NF)	(NF)	(NF)
F-C	712.11	68.29	326.29	269.77	125.80	1502.26
F-B	714.94	21.77	145.60	149.95	56.60	1088.86
Inf-C	722.06	19.78	79.13	178.04	108.80	1107.81
Inf-B	369.86	6.60	61.14	81.53	16.19	535.32
Sum 4 (NF)	2518.97	116.44	612.16	679.29	307.39	

Each of the tables above has two Sum columns, vertical and horizontal. The vertical Sums 1 and 2 represent the total number of MDMs produced, regardless of semantic types, in each NS and learner corpus. These two sums illustrate to what extent the register differences among the NS corpora and the English level among the learner corpora affect the total quantity of produced MDMs. The horizontal Sums 3 and 4 show the total MDMs produced, sensitive to semantic types, in the four NS/learner corpora within each semantic category. Sum 3 shows the quantity of different semantic types of MDM produced by NS, whereas Sum 4 shows how semantic type affects learners' choice of MDMs.

The four corpora in Table 4.32 are ranked in order of relative formality, from the most formal written corpus (BNC-Acad) to the least formal spoken corpus (CANCODE). The four learner corpora in Table 4.33 are similarly presented, from the most formal written corpus, which archives advanced learners' writing (F-C), to the least formal written corpus by intermediate learners (Inf-B).

The values in Sums 1 and 2 again show learners' overuse of MDMs. The number of MDMs in Sum 1 seemed to correlate with the formality of the context in that a greater NF of MDMs often accompanied a more formal register. Although the CANCODE corpus had more MDMs than the Enronsent2 corpus, this might be simply due to the fact that out of the 243 instances of Concessive MDMs in the CANCODE corpus, the very informal MDM *but then again* appeared with a very high frequency: 194 times. In Sum 2, we see a similar trend even though the numbers in the F-B and the Inf-C corpus were very close. In general, the more advanced learners employed more MDMs.

While Sums 1 and 2 showed how formality is related to the amount of MDMs used, Sums 3 and 4 revealed the effect that different semantic types might have on NS and learners' choices of MDMs. Ranking the NF in Sum 3 (NS data) from high to low gives rise to the following order: Additive, Concessive, Appositive, Causal, and Contrastive. At face value, it showed that NS tended to use MDMs in the first three categories and the use of Causal and Contrastive MDMs was less common. Such a tendency is particularly salient in the less formal or spoken registers. Although I have not confirmed this hypothesis, it seems intuitively likely that the less frequent use of Causal and Contrastive MDMs might be due to the Causal and Contrastive single-word markers *so* and *but*, which are both high-frequency words.

Ranking the NF in Sum 4 from high to low, we have a similar order: Additive, Concessive, Causal, Contrastive, and Appositive. Like NS, learners also produced the most Additive and Concessive MDMs; however, learners produced the fewest instances of Appositive MDMs. That learners also produce the most Additive and Concessive MDMs, however, does not necessarily suggest that they have a firm grip on those two semantic categories. The formal-informal ratios in NS and learners' Additive MDMs data (Section 4.3.1) showed that C- and B-level learners in the formal register used Additive MDMs in a native-like manner despite the overuse phenomenon. When writing in the informal register, C-level learners still produced many formal Additive MDMs, but B-level learners did not. Although the F-C and the F-B corpus did not show similar formal-informal ratios to that of the BNC-Acad corpus in the Concessive category, both the two learner groups and NS writers showed a striking resemblance in the proportion of *formal* Concessive MDMs that they produced in the formal

register. The two groups of learners, however, used a greater number of formal Concessive MDMs in the informal register. In other words, using Concessive MDMs in the informal register remains a greater problem for L2 learners.

The comparison of Sums 3 and 4 also reveal that NS and learners have a very different usage pattern in Appositive MDMs. While learners showed an overwhelming overuse of MDMs in most of the semantic categories, a noticeable usage difference was learners' underuse of Appositive MDMs, which was ranked as the third most produced type of MDMs in the NS data. The formal-informal ratio information (see Section 4.3.2) indicated that the two learner groups showed no difference in using Appositive MDMs in the formal and informal register. This suggests that learners might not have noticed the formality differences when using Appositive MDMs.

NS frequently employed Additive and Concessive MDMs, and they produced the fewest Contrastive MDMs. Learners also produced the most Additive and Concessive MDMs but they evidently underused Appositive MDMs. In other words, this frequency difference suggests a different order of acquisition among the five types of MDM between the NS and L2 learners. This is not to say, however, that the most frequently occurring type of MDM in the NS data (Additive) suggests the easiest type to acquire for L2 learners, nor that the least frequent type of MDM (Contrastive) is the most difficult to acquire. For L2 learners, the concept of the Appositive tie might be the most challenging. On the basis of the evidence examined thus far, a tentative conclusion is that learners face the greatest challenge in using Appositive and Concessive MDMs.

4.4.2 The formality makeup of the MDMs

In this section, I first eliminate the semantic boundaries of the MDMs retrieved by merging the five semantic categories and present the MDMs in accordance with the three general formality levels, i.e., formal, neutral, and informal. The aim is to compare the formality composition between the NS and learner corpora. I will then integrate the formality and semantic information of the MDMs produced by learners (see Table 4.35). The NF of MDMs by formality in the eight corpora is shown in Table 4.34.

Table 4.34

NF of MDMs by formality

Corpus	Formal MDMs	Neutral MDMs	Informal MDMs	Sum
BNC-Acad	373.16	194.78	108.95	676.89
BASE	132.35	83.45	334.66	550.46
Enronsent2	138.1	23.21	67.38	228.69
CANCODE	13.31	54.86	317.64	385.81
F-C	798.55	325.61	378.09	1502.25
F-B	496.29	194.93	397.62	1088.84
Inf-C	474.79	227.5	405.54	1107.83
Inf-B	121.1	97.72	316.52	535.34

The NF information that shows the formality makeup in the four NS corpora reveals the formality dynamics in various registers. While the NS in the BNC-Acad

corpus (formal, written English) used the most formal MDMs, NS in the CANCODE corpus (informal, spoken English) produced the most informal MDMs. Although the BASE corpus had its academic attribute, more informal MDMs were found in this corpus. NS in the Enronsent2 e-mail corpus showed the nature of e-mail communication by using the fewest instances of MDMs in general, as using single-word markers might be more efficient.

The formality dynamics in the learner corpora will have to be examined by taking into account learners' English levels. C- and B-level learners produced nearly the same number of informal MDMs in the F-C (NF: 378.09) and the F-B corpus (NF: 397.62). The main differences lay in the category of formal and neutral MDMs, in which C-level learners evidently used more instances of these two types of MDM. C-level learners still used more formal MDMs when writing in the informal register, supported by the NF of the formal MDMs of 474.79. Although B-level learners in the Inf-B corpus produced the fewest instances of MDMs, a substantial portion of these were informal MDMs (NF: 316.52).

The NF information in Table 4.34 was converted to percentages, shown in Figure 4.17. The three different colour bars provide a visual representation of the formality trend in the eight corpora. The bars in the four NS corpora supported the observation on formality dynamics that I discussed earlier. The bars in the four learner corpora showed a clear tendency that learners' English level affected the formality type of MDMs learners produced. While the percentage in the blue bars (formal MDMs) declined from the F-C corpus to the Inf-B corpus, the percentage in the red bars (informal MDMs) increased from 25.17% in the F-C corpus to 59.13% in the Inf-B

corpus. Although C-level learners have shown a preference for using formal MDMs in the informal register (Inf-C), the overall trend showed that L2 learners' English level influenced their choices in both the quantity and the formality-types of MDMs.

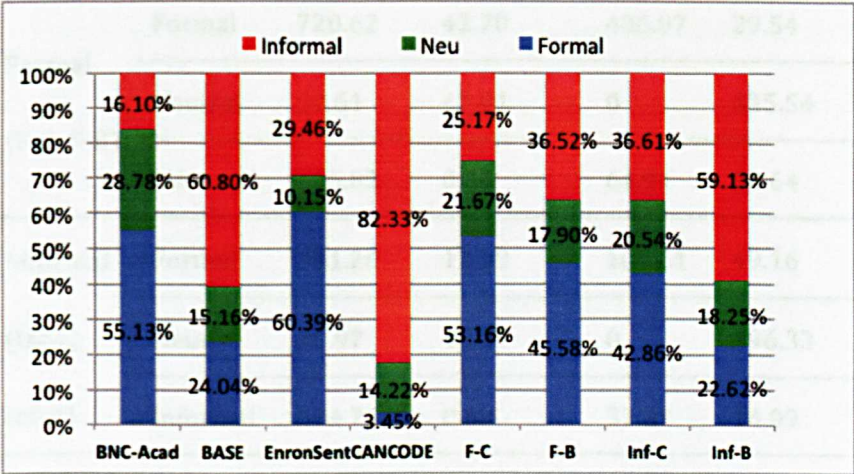


Figure 4.17 Formality makeup of the MDMs produced

In order to see the effect that formality and semantics have upon CLC learners' choices of MDMs, in what follows I look at the frequency information of the five semantic categories from two registers: formal and informal. The formal register includes data retrieved from both the F-C and the F-B corpus, whereas the informal register includes information from the Inf-C and the Inf-B corpus. Table 4.35 presents a breakdown of the formality and semantic information.

Table 4.35

The breakdown of learners' MDMs by formality and semantics

Register	MDMs	Additive (NF)	Appositive (NF)	Causal (NF)	Concessive (NF)	Contrastive (NF)
Formal (F-C, F-B)	Formal	720.62	42.70	406.97	29.54	95.01
	Neutral	50.61	47.01	0	335.54	87.38
	Informal	655.82	0.34	64.91	54.64	0
Informal (Inf-C, Inf-B)	Formal	381.26	12.89	107.61	49.16	44.97
	Neutral	35.97	12.89	0	196.33	80.03
	Informal	674.70	0.60	32.67	14.09	0

The earlier discussion showed that both Additive and Concessive MDMs occurred frequently in the four learner corpora. The breakdown of Additive and Concessive MDMs in Table 4.35 showed clearly the distribution of these MDMs in the formal and informal register. Half of the total informal Additive MDMs that learners used occurred in the formal register (NF: 655.82), while the other half were in the informal register (NF: 674.7). Of all the 1101.88 instances of formal Additive MDMs, only 35 percent (NF: 381.26) occurred in the informal register. In other words, while the CLC learners used more formal Additive MDMs in the formal register, their use of informal Additive MDMs was similar in the two registers. The two distributions showed that learners were more capable of using formal Additive MDMs in the formal register.

In the case of Concessive MDMs, the neutral Concessive MDMs occurred frequently in both the formal and the informal register. More formal Concessive MDMs

were found in the informal register, and more informal Concessive MDMs were observed in the formal register. The contextual mismatch in Concessive MDMs suggests that a greater usage problem lies in learners' insensitivity to register difference. Another similar problem was seen in Appositive MDMs. Learners produced equally small numbers of both formal and informal Appositive MDMs in the two registers, indicating L2 learners' problematic use of Appositive MDMs in accordance with register difference.

4.5 Summary and conclusion

In this chapter, I have examined how learners employed various MDMs in the formal and informal register and compared learner data with the NS data. By creating several learner sub-corpora on the basis of learners' English levels (C2, C1, B2, B1) and the registers (formal and informal), I have examined learners' use of MDMs from two perspectives—formality and semantics. A relationship was found to exist between the level of English proficiency and the quantity and types of MDMs that learners produced. Although the corpus evidence suggested a high correlation between English level and the quantity of MDMs learners used, this is not to say that the MDMs produced by advanced learners were less problematic. A close examination of the formality and the semantics of MDMs showed that while both advanced (C-level) learners and intermediate (B-level) learners noticed the formality in the formal register and used MDMs accordingly, C-level learners also produced more formal MDMs in the informal register. As the CLC corpus archives learners' examination scripts, the fact that the writing is in the exam format might explain C-level learners'

preference for formal MDMs in the informal register. Another plausible reason might be that C-level learners have been in the habit of writing formally and such practice has inevitably affected their linguistic choices in the informal register.

The findings reported thus far provide answers to the second research question raised in Chapter 1:

RQ2. Provided that an MDM formality continuum does exist, do learners perceive the formality differences in MDMs and use these markers accordingly? Or do L2 learners show a similar lack of register awareness in using MDMs, as is the case with SDMs?

The answer is that the CLC data showed that L2 learners paid the most attention to the formality difference in Causal MDMs. While they used formal Additive MDMs in formal discourse accordingly, they still used more formal MDMs in the informal register. Employing both Concessive and Appositive MDMs in accordance with the contextual formality seems to present a major challenge to the L2 learner.

This discussion has also shown that interpreting the quantity of MDMs has to be made by taking into account both the formality factor and the semantic factor. The analysis in the semantics category of MDMs indicated that some types of MDM tended to occur more frequently than others in both the NS and the learner corpora. For example, while learners produced a large number of Additive and Concessive MDMs, as is the case in the NS data, learners used Appositive MDMs in a very limited way. When considering the formality types within the Additive and Concessive categories, I discovered that learners faced a greater problem in using Concessive MDMs in

accordance with the formality of the context. In marked contrast to NS' frequent usage of Appositive MDMs, learners' use of Appositive MDMs was quite limited. The reason that lies behind this avoidance will be further explored in Chapter 6.

In the next chapter, I present two studies that are designed to test L2 learners' pragmatic awareness in using MDMs in the formal and informal registers. The purpose is to see whether the corpus-based findings reported thus far will be supported by experimental results. Another purpose is to see if the devised experiments are suitable for testing learners' pragmatic awareness.

Chapter 5 - Reflecting L2 pragmatic awareness in two registers—an experimental approach using corpus-informed material

I have shown in Chapter 4 that L2 learners show different patterns in their pragmatic awareness of the formal and informal register, depending on the semantic types of the MDMs examined. I discovered that the CLC learners showed pragmatic awareness in the use of Additive MDMs, evidenced by their higher proportion of formal Additive MDMs used in the formal register. In using Causal MDMs, however, the CLC learners were found to use formal forms in both the formal and the informal register. The tendency to use more formal Concessive MDMs in the informal register suggests a lack of pragmatic awareness when learners write in informal style. Learners were found to display the fewest problems when using Contrastive MDMs in both the formal and the informal register. The neutral Concessive and Contrastive MDM, *on the other hand*, was the ubiquitous phrase employed across the four groups of learners in both the formal and informal registers.

Of the five types of MDM, Appositive MDMs were found to occur the least frequently in the CLC data. The formality analysis in this category also suggests that only the C-level learners employed the Appositive MDMs in the formal register. The insufficient representation of Appositive MDMs among other learners might be indicative of an avoidance strategy. Whether or not it is an avoidance strategy that leads to the underuse phenomenon, the fact that both the C- and the B-level learners used Appositive MDMs in a rather restricted manner does imply that learners face a greater level of difficulty. The mixed evidence in Chapter 4 suggests that L2 learners

are not completely insensitive to register difference. Rather, learners' (in)sensitivity to register is influenced by the semantic type of MDM.

As spoken learner corpora are still underrepresented, the literature reported in Chapter 2 has mainly examined L2 writing and reported problems with DM use in written discourse. Although I examined the CLC data in both the formal and the informal register, the examined data was all from a database of written, not spoken, texts. The CLC writing stored in the informal register is by no means equal to the conversational type of discourse produced by L2 learners. One neglected area in researching learners' pragmatic awareness in registers, therefore, is how L2 learners use MDMs in spoken discourse. Further, since linguistic awareness is the individual's cognitive reflection of the context or the environment, another intriguing question is whether the corpus findings of L2 learners' lack of register sensitivity in using MDMs are supported by any psycholinguistic evidence.

Pragmatic awareness is embodied in competence, which "one cannot directly measure..." (Canale and Swain, 1980, p. 6) and one possible way to observe this competence is via individuals' performance. The pragmatic awareness tested in this chapter, as a result, was observed via two studies. The traditional format in testing language users' knowledge of linking DMs is usually comprised of two sentences that are juxtaposed. Sentence (1) shows an example of such representation:

(1) The weather was not very nice. _____, John went to see a movie.

Possible phrases to fill in the blank include Causal MDMs, *as a result* or *because of that*, and Concessive MDMs, *despite that* or *but then again*. Employing a Causal marker suggests that an outdoor activity (a picnic or a trek) was originally planned, but was cancelled. Inserting a Concessive marker, on the other hand, suggests that John went out regardless of the poor weather conditions. The adjacency pair in (1) shows that more than one possible inference arises when insufficient background information is provided. Such design would undermine the purpose of the two studies, which is to lead readers to only one possible coherence relation so that the formality comes into play. Additionally, the limited information in an adjacency pair does not prepare readers to identify whether or not a passage is written with a formal or informal style. As a result, the two studies included passages instead of adjacency pairs as the testing materials. In addition, unlike the conventional laboratory experiments, which use designed material, I took a new approach by including in the two studies corpus-informed, authentic data.

Two techniques were incorporated into each of the studies: formality mismatch and register comparison. Formality mismatch is created when the formality of a context and that of an MDM do not match. For example, when the informal Concessive MDM *but then again* is used to take place of the formal Concessive MDM *in spite of that* in a passage written with complex structure and formal wording, a formality inconsistency is created. This inconsistency is referred to as a formality mismatch. The second technique included in both studies involved the two registers compared in testing language users' pragmatic awareness. Study 1 compared the formal written and the informal spoken register, while Study 2 compared formal and informal written

discourse. With the purposely created formality mismatch in the two studies, I aim to explore the extent to which the two designs reflect the pragmatic awareness of language users, and further, to see if the corpus data in Chapter 4 can be verified through the study results.

The two studies included a reaction time task and a multiple-choice discourse completion task (MDCT) respectively. Study 1 involved a self-paced reading task, which recorded the timed duration of each participant's reading of each sentence. When reading a passage in which a formality-inconsistency device is embedded, I hypothesize that the reading time of the target sentence (the sentence where the MDM occurs) could be used as an indicator of the language users' pragmatic awareness. One of the hypothesis scenarios is that language users who have developed a pragmatic awareness of register difference would notice the formality mismatch and that this will produce a prolonged reading time. The other scenario is that language users do not notice the formality mismatch and produce the same length of reading time, when compared with the time spent in reading formality-matched sentences.

Study 2, which compared formal and informal written discourse, was designed as a multiple-choice discourse completion task (MDCT). Participants were asked to read ten passages in which the linking marker was missing, and to choose an appropriate MDM from the options provided. The ten MDMs provided (in five pairs) were from five semantic categories, chosen on the basis of the corpus data reported in Chapter 3. Each pair represented one semantic category that included a very formal and a very informal MDM. For example, the pair for the Causal category was *as a result* and *because of that*. The assumption is that when a competent language user reads a

formal passage in which the Causal marker is missing, that user will accommodate to the formality and will therefore choose the formal Causal MDM *as a result* instead of the synonymous, informal counterpart *because of that*.

Each of these designs is a new attempt to observe learners' awareness of or sensitivity to register differences. The discussion of the experiment results will focus on the observed pragmatic awareness, and in addition, on the plausibility of the two methodologies in testing learners' pragmatic awareness.

5.1 Theoretical framework

The theoretical frameworks used in the two studies are the noticing hypothesis and accommodation theory. I have given a detailed account of the two theories in Chapter 2. In what follows, I elaborate further on how the two frameworks were used in analysing and evaluating the two study results.

5.1.1 Noticing hypothesis

Schmidt (1990, 1993) separated *noticing* from *understanding*, with the former being used to describe attention that is paid to the more concrete phenomena, such as item learning, and the latter describing attention paid to the more abstract level of meaning, rules, or generalization. I have adopted these two types of attention in describing the attention flow of the participants in the two studies. Noticing was used to describe participants' attention to different semantic types of MDM. For instance, when a participant reads a short passage in which the linking Causal MDM is missing, this participant shows his noticing by choosing MDMs like *as a result* or *as a*

consequence, rather than other semantic types of MDM. Understanding, on the other hand, is used to describe whether or not participants have taken into account the formality of the context and choose an MDM that is at the same level of formality. For example, when reading a formal passage in which the Additive MDM is missing, participants with fully developed awareness would pay attention to the formality of the context and choose, in this case, *in addition*. On the other hand, when participants only notice the semantic differences of the MDMs and choose the informal Additive MDM *on top of that* for this formal passage, such decision shows that the awareness of register difference is not yet fully developed.

5.1.2 Accommodation theory

Accommodation theory is used to account for the linguistic choices that a speaker makes in order to decrease (converge), or to increase (diverge) the distance between him/herself and an interlocutor. Although the two studies in this chapter were neither speech-based nor face-to-face interactions, I include accommodation theory in analysing the data and the rationale is given below.

For accommodation to take place, one premise is that language speakers notice or perceive the linguistic differences and then make a decision as to whether or not to accommodate. The concept of perceiving the linguistic differences used in accommodation theory can be applied to the formality differences that I explored in the two studies. Language users with well-developed pragmatic awareness would accommodate and *converge* in the reading task; that is, they will expect low-frequency words/phrases when reading formal passages, and at the same time, they will expect a

more colloquial type of wording when reading informal passages. Such convergence will accordingly lead to their choice of appropriate MDMs.

Both of the studies involved a reading task. The nature of reading activity gives the impression of a solitary process. The reading tasks, however, were manipulated in such a way that the accommodative process had to be enforced or activated in order for the task to be completed successfully. As mentioned earlier, a formality inconsistency was devised in both of the experiments. In Study 1, it was hypothesised that participants who adjusted to the formal/informal tone in the reading task would be more likely to slow down their reading pace when encountering the planned formality mismatch. In the MDCT task where participants were asked to choose a phrase for the missing MDMs, participants who adjusted to the formal/informal tone would choose appropriate MDMs that were at the same formality level as the one suggested in the context. In other words, accommodation theory was used to describe how competent language users adjusted to the formality level suggested in the context as well as the linguistic choices that followed. Such accommodation facilitated the noticing of the formality mismatch planned in Study 1 and the selecting of the missing MDMs in Study 2.

As accommodation theory includes two directions—convergence or divergence— participants who did not accommodate to the formality of the context should not be seen as making a divergent choice. All the participants were aware of their involvement in the two studies and they were asked to perform in accordance with the instructions given. The participants were told to read as fast as possible (Study 1) or to choose the most natural phrase for the missing blanks (Study 2).

Participants who neither accommodated to the formality level nor made appropriate choices of MDMs were assumed to be oblivious to the formality differences. Another explanation of participants' unawareness of the formality differences might be the study designs. I will elaborate more on these considerations in the discussion section.

5.2 The two registers

I have written a few paragraphs in Section 3.1 (Chapter 3) on the factors and criteria that define a register, and I have reviewed relevant literature that addresses differences among various registers. Of the cited studies, Finegan's (2008) discussion of registers was used in this chapter. Finegan considered registers as a "situational continuum" at the two ends of which sit formality and informality (2008, p. 319). I followed Finegan's description and used the term *registers* in a broad sense, so that the two registers or styles discussed in this chapter are formal academic discourse on the one hand, and very informal or colloquial discourse on the other. Specifically, the two registers compared in Study 1 were formal written discourse (BNC-Acad) and informal spoken discourse (CANCODE). Making a direct comparison between written and spoken language in a reaction time task, however, resulted in problems in data interpretation (see Section 5.3.6.7). To remedy this flaw, the two registers that I compared in Study 2 were formal written discourse (BNC-Acad) and informal written discourse (Enronsent2 e-mail corpus).

5.3 Study 1: a reaction time design

I designed a self-paced reading task by using the reaction time experimental paradigm on E-prime. The task was designed on the basis of the following scenario. Imagine that two different types of text are presented to a reader. The two irrelevant short texts are written with very different styles. While one, referred to as Text A, is written in a very academic and formal style (for example, a research report about the human brain), the other, named Text B, is composed in a very colloquial and informal style (a spoken discourse segment about a carnival experience). One feature that both Texts A and B share is the Additive MDMs, with *in addition* employed in Text A and *what's more* used in Text B, as (2) and (3) show below.

(2) *In addition*, the world is geared towards logical thinking. (BNC-Acad)

(3) Erm. *What's more*, not just that it's doing the kid's eardrums damage but also they'll use the child as a battering ram.

(CANCODE)

Although the rest of the texts are not shown here, it is evident that the formality levels of the MDMs and the contextual cues are coherent. Sentences (2) and (3) can be referred to as the formality-consistent examples. When *what's more* is interchanged with *in addition*, we have a formality-inconsistent case, as (4) shows.

(4) Erm. *In addition*, not just that it's doing the kid's eardrums damage but also they'll use the child as a battering ram.

Reading (4) alone might not incur a strong sense of formality-inconsistency; yet, it is likely that by reading the whole passage in which (4) is embedded, participants would perceive the planned formality mismatch. I hypothesize that participants with well-developed pragmatic awareness of register differences would accommodate in the reading process and notice the formality mismatch, which in turn incurs a sense of oddness. This impression of oddness would slow down the reading pace, and as a result, produce a longer reading time.

5.3.1 Participants

Eighteen native speakers of English (NS) and another eighteen L2 learners were included in the first study. All of the L2 learners were PhD students at the University of Nottingham, with various L1 backgrounds. By the time they took part in this study, these L2 learners had spent an average of 1.5 years in the UK and the average IELTS score was 7.0. This group of L2 learners was therefore considered an intermediate to advanced group. That being said, data of one of the L2 learners was removed from the analysis because he spent relatively longer time reading the texts. In the end, data retrieved from the eighteen NS and seventeen L2 participants were analysed.

5.3.2 Materials

On the basis of the formality continuum reported in Chapter 3, six target MDMs in three semantic categories were included, and each was controlled in terms of length

and syllables (see Table 5.1). Another three MDMs that tend to be neutral in terms of formality were also included to serve as control items. The reason why only three semantic categories (six MDMs) were included was because of the length of the study. Although there were only six targeted MDMs, adding the filler items and another three neutral MDMs from the three semantic categories made the reading task a long study. Considering the cognitive resources that one can dedicate to the task at hand, i.e., the time-on-task effect, I therefore only included three semantic categories so that participants could concentrate on the task.

Table 5.1

A summary of the six critical items

Semantic category	Register	MDMs	Syllables	Length
Additive	spoken	<i>what's more</i>	2	9 letters
	written	<i>in addition</i>	3	10 letters
	neutral	<i>besides that</i>	3	11 letters
Causal	spoken	<i>because of that</i>	4	13 letters
	written	<i>in consequence</i>	4	13 letters
	neutral	<i>for that reason</i>	4	13 letters
Concessive	spoken	<i>but then again</i>	4	12 letters
	written	<i>despite that</i>	3	11 letters
	neutral	<i>that being said</i>	3	13 letters

To find suitable passages for this study, I searched for the six target MDMs in the CANCODE corpus (Carter & McCarthy, 1995) and the written section of British National Corpus (BNC-W). The three neutral MDMs were searched in the British Academic Written English Corpus (BAWE). The CANCODE corpus provides informal, spoken data, whereas BNC-W provides formal, written materials. Even though the BAWE corpus includes both NS and non-native speaker (NNS) student writing, I purposely checked the metadata (e.g. text type, genre, L1 background, etc.) and selected only passages written by NS of English. The reason for choosing the BAWE corpus is the nature of the corpus, which is academic-oriented, and yet not as formal as the formal written section in the BNC corpus. Only passages that included MDMs appearing in the sentence-initial position were considered potential materials. After checking the grammar as well as the wording, I selected nine reading passages: six for the target MDMs and three for the neutral MDMs (See Appendix 2). The six target passages had two versions: a formality-consistent version (texts with the original MDMs) and a formality-inconsistent version (texts with MDMs being replaced, as shown in sentence (4) previously).

In order not to reveal to the participants the purpose of the task, I created six experimental lists (see Table 5.2). Each list had one Additive text, one Causal text, one Concessive text, and twelve filler texts⁵ (see Appendix 3 for Experimental list 1). The twelve filler texts were retrieved from the British Academic Written English Corpus (BAWE) and the CANCODE corpus. The criterion for selecting the neutral passages as well as the filler passages was that both formal and informal types of text were

⁵ The twelve fillers were short texts with MDMs that were not targeted, e.g. *to begin with*, *for one thing* and *in other words*, etc.

included. Each list included fifteen texts and each text contained approximately 114 to 138 words. The conditions in Study 1 were counterbalanced in a Latin Square design so that participants would not read the same text, nor encounter the same MDM. The carry-over effect was therefore avoided. Also, the order of the texts was randomized.

Table 5.2

The six experimental list (Neu: neutral, F: formal, Inf: informal. FC: formal context, InfC: informal context)

List 1	Inf Additive <i>what's more</i> in InfC	F Causal <i>in consequence</i> in FC	Neu Concessive <i>that being said</i>	12 fillers....
List 2	F Additive <i>in addition</i> in InfC	Inf Causal <i>because of that</i> in FC	Neu Concessive <i>that being said</i>	12 fillers....
List 3	Neu Additive <i>besides that</i>	Inf Causal <i>because of that</i> in InfC	F Concessive <i>despite that</i> in FC	12 fillers....
List 4	Neu Additive <i>besides that</i>	F Causal <i>in consequence</i> in InfC	Inf Concessive <i>but then again</i> in FC	12 fillers....
List 5	F Additive <i>in addition</i> in FC	Neu Causal <i>for that reason</i>	Inf Concessive <i>but then again</i> in InfC	12 fillers....
List 6	Inf Additive <i>what's more</i> in FC	Neu Causal <i>for that reason</i>	F Concessive <i>despite that</i> in InfC	12 fillers....

5.3.3 Procedure

The study took place in a linguistic lab. The presentation of the passages was done via the E-prime reaction time paradigm. Each passage was shown within seven slides on a computer monitor, with each slide including one or two sentences. Participants pressed a response key to read the next slide and the time taken to press the keys between each slide was recorded. A closure slide appeared right after the seventh slide that informed the participants that it was the end of the passage and a True/False question would pop up once they pressed the response key. On the ninth slide, participants read a simple True/False comprehension question. Participants were told to read as quickly as possible and were given a practice block before they began the task. All the participants volunteered to take part in this experiment. Three NS and three L2 learners were allocated to each experimental list. In other words, neither the eighteen NS nor the seventeen L2 participants read all of the materials.

5.3.4 Measures

The total time taken to read the target sentences (where MDMs were embedded) was divided by the total number of syllables of the target sentences. That is, the analysis was run on the basis of the mean reaction time (RT) in milliseconds per syllable. The rationale is that the length of each target sentence varies and it is reported that the syllables of words do affect the reaction time (Zelinsky & Murphy, 2000; New et al., 2006).

This study comprised a 6x2 factorial design. The two independent variables are CLASS (six levels, formal Additive, informal Additive, formal Causal, informal Causal, formal Concessive, informal Concessive) and REGISTER (two levels, formal and

informal]). The statistical methods used in analysing the data include paired samples t-tests and two-way ANOVA.

5.3.5 A common ground for comparing spoken with written language

A direct comparison between the written and spoken texts on the basis of reading time duration might sound unjustified, as the two types of text are essentially different. Intuitively, reading academic writing takes a longer time. A common ground is therefore needed to serve the comparison purpose. Literature on spoken and written languages has been focusing on the differences observed from a collection of language data (Halliday, 1985; Chafe, 1985; Stubbs, 1980; Biber, 1988), and various parameters such as syntactic constructions, morphological forms, vocabulary, etc., are set so that one differs from the other. Balkin (1994) made a straightforward point in the comparison of speech and writing: "we cannot efface the distinction between speech and writing because they are not in fact identical in all respects" (p.398). To argue or to further illustrate the differences between the two is by no means within the scope of this study. Nevertheless, including both types of language mode is a prerequisite for the *formality mismatch* technique to be effective. If informal speech and formal writing are dissimilar in nature, the direct comparison shown in the experimental design does not seem comparable, as any measured difference in reaction time might be simply due to the differences in the two modes.

To solve this problem, and to provide a common ground for comparison, I considered various methods of measurement, such as type/token ratio and Lexile, a readability analyser. Type/token ratio, or lexical density, is used to see how

linguistically complex a file or text is. Applying type/token ratio to text analysis, however, needs to be done with some caution. Comparing written and spoken discourse from the FLOB corpus and the BNC corpus with the type/token ratio, Baker (2006) maintained that "written language tends to contain a higher proportion of unique words, whereas informal spoken language is more lexically repetitive" (p. 52). He further suggested that the ratio tends to be useful when a small text is examined, defining 'small' as under 5000 words. In line with Baker's claim, Reed (2000) commented on calculating lexical density, and stated that "the figure obtained for some of the statistics varies according to the length of the text" (p. 201). Richard et al. (1987) also concluded that when applying type/token ratio to a text consisting of less than 200 words, the analysis does not yield reliable statistics on lexical diversity. Hess et al. (1986) reported a similar result in their analysis of children's speech. They reported that a text length of 50 to 100 words is not of sufficient length to obtain a reliable type/token ratio. On the basis of these findings, it seems that when applying the type/token ratio in analysing lexical density, the best range of text length is between 200 and 5000 words. As the length of the short passages used in this task is between 110 and 140 words, type/token ratio was therefore not considered. Alternatively, I adopted Lexile (Lexile Framework © for Reading 2008), an online text readability and difficulty analyser, in seeking the common ground for the two different types of texts used.

Research has utilized Lexile in fields such as education and health care in investigating comprehension or literacy-related issues (Stenner et al., 2007; Davis et al., 2006; Williamson, 2008). A Lexile measure indicates the level of readability or

difficulty of a text. On the basis of a 600-million word database, Lexile takes into account the semantic complexity, word frequency, and sentence length in measuring the readability of a text. The possible values range from below 200L to 1700L. According to a scholastic professional paper on Lexile (Lexile, 2008), "... a simple picture book might have a Lexile measure of 100L, while a college textbook might be measured at 1700L or higher" (p.5).

This readability can in turn be seen as the difficulty level of a text. All the study's passages in the twelve conditions (the formality-consistent, formality-inconsistent and neutral cases) were uploaded to Lexile for analysis. Table 5.3 gives an overview of the analysis result.

A first look at the table gives the impression that the formal and informal Additive and Causal passages show a notable difference (Additive FMFR: 1230L vs. Additive InfMInfR: 740L; Causal FMFR: 1390L vs. Causal InfMInfR: 840L); the two Concessive passages, on the other hand, do not show a marked difference (FMFR: 1420L vs. InfMInfR: 1300L). Despite the discrepancies observed, these Lexile measures were used in interpreting the results and the discussion, shown in Section 5.3.6.4.

Table 5.3

Result of Lexile analysis of the twelve conditions (FM-FR: formal MDM in formal register, FM-InfR: formal MDM in informal register, InfM-FR: Informal MDM in formal register, InfM-InfR: informal MDM in informal register)

	Word count	Mean Log Word Ferq.	Mean Sentence Length	Lexile measure
Additive MDM (in addition vs. what's more)				
FM-FR	126	3.21	18.00	1230L
FM-InfR	114	3.80	12.67	740L
InfM-FR	126	3.24	18.00	1220L
InfM-InfR	114	3.80	12.67	740L
Causal MDM (in consequence vs. because of that)				
FM-FR	133	3.54	26.60	1390L
FM-InfR	119	3.85	14.88	840L
InfM-FR	134	3.56	33.50	1550L
InfM-InfR	120	3.87	15.00	840L
Concessive MDM (despite that vs. but then again)				
FM-FR	137	3.17	22.83	1420L
FM-InfR	114	3.90	28.50	1300L
InfM-FR	138	3.19	23.00	1420L
InfM-InfR	115	3.91	28.75	1300L

5.3.6 Results and Discussion

I present the results by giving an overview of the NS and the L2 participants' mean RT in the twelve conditions, followed by the comparisons of the RT in the formality-consistent and formality-inconsistent conditions. Statistical results of the pair-sampled t-tests and two-way ANOVA will be presented subsequently.

5.3.6.1 The mean RT in the twelve conditions

The mean RT of the L2 and NS participants in the twelve conditions are shown in Table 5.4. The formality-inconsistent conditions are marked in the first column with a star sign. Passages that are identical except for the linking MDMs are paired together. For example, condition 1 (FAdd-FC) and 2 (InfAdd-FC) in Table 5.4 are in fact the same passage written in formal tone and wording, except that the former includes the formal Additive MDM *in addition*, whereas the latter has the informal Additive MDM *what's more*. That is, the former is a formality-consistent condition, whereas the latter is a formality-inconsistent condition.

One easy calculation is to find out whether there is any RT difference between the two conditions. Take again conditions 1 and 2, for example. The NS participants' mean RTs in the two conditions were 116.2 and 224.027 milliseconds (ms). That is, we observe a longer responding time in the formality-inconsistent condition. The difference between the two RT, 107.827ms, suggests a level of pragmatic awareness, or the NS participants' noticing of the formality mismatch. The L2 participants' mean RTs in the first two conditions, on the other hand, are 181.937 and 182.563ms. The mean RT in the formality-inconsistent case is only slightly longer, resulting in the very

small RT difference of 0.626ms. This seems to suggest that the L2 participants might read the two passages without noticing the formality mismatch.

Table 5.4

Summary of the NS and L2 participants' mean RT in the 12 conditions

Conditions	L2 Mean RT (ms)	NS Mean (ms)
1. FAdd-FC	181.937	116.200
2. InfAdd-FC*	182.563	224.027
3. FAdd-InfC*	160.810	208.347
4. InfAdd-InfC	281.023	186.847
5. FCau-FC	198.307	151.193
6. InfCau-FC*	206.343	140.957
7. FCau-InfC*	214.045	255.427
8. InfCau-InfC	226.637	157.243
9. FCon-FC	181.063	94.950
10. InfCon-FC*	139.780	162.447
11. FCon-InfC*	211.443	222.610
12. InfCon-InfC	270.127	409.640

Following the same mechanism, I grouped together the conditions that used the same passages but differed from each other in the embedded MDMs and calculated the RT differences in Table 5.5.

Table 5.5

Summary of RT differences in each paired condition

Paired conditions	L2 mean RT (ms.)	NS mean RT (ms.)
1. InfAdd-FC* / FAdd-FC	0.626	107.827
2. FAdd-InfC* / FAdd-InfC	-120.213	21.5
3. InfCau-FC* / FCau-FC	8.036	-10.236
4. FCau-InfC* / InfCaul-InfC	-12.592	98.184
5. InfCon-FC* / FCon-FC	-41.283	67.497
6. FCon-InfC* / InfCon-InfC	-58.684	-187.03

There are six pairs in Table 5.5. The only pair where L2 participants showed a noticing of the formality mismatch is when they read the formal passage for Causal MDMs (pair 3). The L2 participants showed a longer RT in reading the formal Causal passage with the informal MDM *because of that*. In the rest of the five groups, L2 participants either displayed a shorter reading time in the formality-inconsistent conditions or showed no difference at all. The NS participants, on the other hand, displayed a longer reading time in four pairs (pair 1, 2, 4, and 5). The most unexpected result is shown in pair 6, where both the L2 and NS participants seemed to have a longer reading time in the formality-consistent condition, which led to negative gains of -58.684 and -187.03 respectively.

Having shown the two groups' RT in the twelve conditions as well as the RT differences, in what follows I present a visual comparison of the two groups' task performance in the two general conditions.

5.3.6.2 Mean RT in formality-consistent conditions

Formality-consistent conditions are passages in which the formality of the MDMs and the formality of the texts are consistent. The assumption is that both the NS and the L2 participants would read at a normal pace. Taking into account that English is a second language for the L2 participants, these participants would presumably have a longer response time even when reading passages in formality-consistent conditions. Figure 5.1 shows the RT (per syllable in milliseconds) of formality-consistent cases.

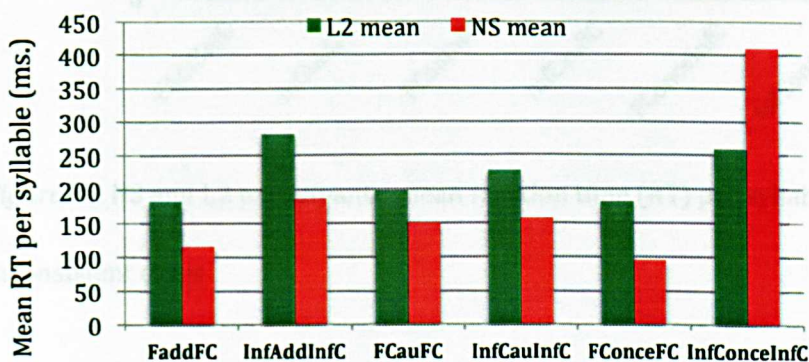


Figure 5.1 Mean reaction time (RT) per syllable in consistent cases

As expected, the mean RTs of L2 participants, shown in the green bars, were longer than those of NS. Unquestionably, the NS participants read faster (hence the shorter red bars) than the L2 participants in a normal reading condition. In the case

where the informal Concessive MDM occurred in an informal context (*InfConceInfC*), however, the NS participants had a longer reaction time.

5.3.6.3 Mean RT in formality-inconsistent conditions

The hypothesis in formality-inconsistent conditions is that participants with well-developed pragmatic awareness might find the target sentences ‘odd’ and they would hesitate in pressing the response key, leading to a longer response time. The mean RT in these conditions is shown in Figure 5.2.

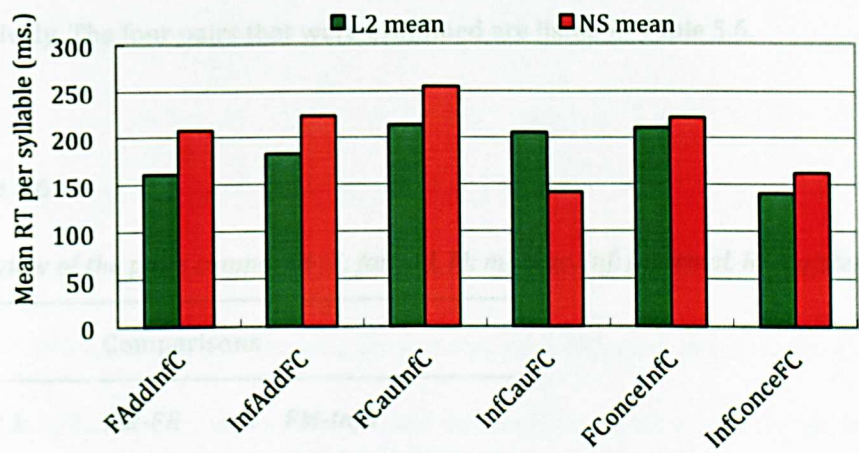


Figure 5.2 NS and L2 participants’ mean reaction time (RT) per syllable in inconsistent cases

The bars in Figure 5.2 show that the NS participants’ reaction time (the red bars) was longer than that of L2 learners (the green bars). In most of the cases, NS spent a longer time reading the formality-inconsistent sentences, with only one exception: the informal passage with which the formal Causal MDM occurred (*InfCauFC*).

5.3.6.4 Paired samples T-tests

Although Study 1 included both NS and L2 participants, a direct comparison of the two types of data via a statistical test, such as an independent T-test, was not considered. The main concern is that because L2 participants tended to have a longer reading time when reading in English, it is not easy to differentiate whether any significant finding is a result of L2 participants' pragmatic awareness, or is due to L2 participants' naturally longer reading behaviour. Instead of running independent T-tests, I conducted two paired samples t-tests to examine the NS and L2 data respectively. The four pairs that were examined are listed in Table 5.6.

Table 5.6

Overview of the pairs compared (F: formal, M: marker, Inf: informal, R: register)

Comparisons		
Pair 1	<i>FM-FR</i>	vs. <i>FM-InfR</i>
Pair 2	<i>InfM-FR</i>	vs. <i>InfM-InfR</i>
Pair 3	<i>FM-FR</i>	vs. <i>InfM-FR</i>
Pair 4	<i>FM-InfR</i>	vs. <i>InfM-InfR</i>

The comparisons made in Pair 1 and Pair 2 show how RT is affected by reading the same MDM in two different contexts, formal and informal. In Pair 3 and Pair 4, the comparisons are made to see how different MDMs are read in the same formal or informal registers. In what follows, I present the NS and L2 data according to the semantic categories of the MDMs.

NS data

Additive MDMs: in addition vs. what's more

The paired-samples t-test in Pair 1 showed that the mean RT of reading *in addition* in the informal context was significantly longer ($M = 208.35, SD = 43.17$) than reading the same MDM in the formal register ($M = 116.2, SD = 29.11$), $t(2) = 9.01, p = .012$. That is, the NS participants displayed a longer reading time when reading the formality-inconsistent condition of *in addition*. The paired-samples t-tests run for the other three pairs did not show any significant differences.

Causal MDMs: in consequence vs. because of that

None of the RT comparisons made in the four pairs was found to be significantly different. That is, the NS participants did not display a different response time, regardless of the conditions (formal or informal register) in which the formal MDM *in consequence* and the informal MDM *because of that* appeared. Nevertheless, a nearly significant p score was found for Pair 1, with $t(2) = 3.91, p = .06$, suggesting that a slightly longer reading time was perceived when the formal Causal MDM *in consequence* occurred in the informal register ($M = 255.42, SD = 89.71$).

Concessive MDMs: despite that vs. but then again

The comparisons in Pair 1 and Pair 2 were found to be significant. In Pair 1, the NS participants showed a longer RT when reading the formal MDM *despite that* in the informal context ($M = 222.61, SD = 28.04$) than when reading it in the formal context

($M = 94.94, SD = 18.48$), $t(2) = 10.808, p = .008$. A similar phenomenon was observed with the informal MDM *but then again* in Pair 2. *But then again* was found to have longer response time in the informal context ($M = 409.64, SD = 75.42$) than in the formal context ($M = 162.44, SD = 57.1, t(2) = 5.03, p = .037$). This is different from all the significant cases reported above, where the significance was found in the formality-inconsistent cases. I will pursue this further in the discussion of the ANOVA analysis.

L2 data

None of the comparisons made in the L2 data was found to be significant. The results of the comparisons are shown in Table 5.7.

Table 5.7

Descriptive statistics of the paired samples T-test in NNS data

		Mean	N	Std.	t-score	Sig.
<i>Additive MDMs</i>		(ms.)		Deviation		(2-tailed)
Pair 1	FMFC --	181.93	3	72.60	.305	.789
	FMInfC	160.81	3	49.90		
Pair 2	InfMFC	182.56	3	55.27	-1.515	.269
	InfFInfC	281.02	3	74.86		
Pair 3	FMFC	181.93	3	72.60	-.009	.994
	InfMFC	182.56	3	55.27		
Pair 4	FMInfC	160.81	3	49.90	-1.699	.231
	InfMInfC	281.02	3	74.86		
<i>Causal MDMs</i>						
Pair 1	FMFC --	198.30	3	53.43	-1.139	.373
	FMInfC	579.70	3	633.53		

Pair 2	InfMFC	206.31	3	52.18	-.498	.668
	InfFInfC	226.64	3	41.83		
Pair 3	FMFC	198.30	3	53.43	-.131	.907
	InfMFC	206.31	3	52.18		
Pair 4	FMInfC	579.70	3	633.53	.963	.437
	InfMInfC	226.64	3	41.83		
<i>Concessive MDMs</i>						
Pair 1	FMFC --	181.06	3	38.34	-1.118	.380
	FMInfC	211.44	3	21.47		
Pair 2	InfMFC	324.99	3	321.09	.363	.751
	InfFInfC	258.13	3	40.77		
Pair 3	FMFC	181.06	3	38.34	-.803	.506
	InfMFC	324.99	3	321.09		
Pair 4	FMInfC	211.44	3	21.47	-1.788	.216
	InfMInfC	258.13	3	40.77		

Discussion

Neither the NS nor L2 participants showed a significant difference in their RT when comparing the formality-inconsistent passages with the formality-consistent passages (Pair 3 and 4). Nevertheless, one interesting finding is shown in Pair 1 and Pair 2 comparisons, in which the same MDM is compared in two different types of text (formal written and informal spoken). As discussed earlier, written and spoken texts are different in nature, so any RT difference observed in reading the two kinds of texts might simply have arisen because of the two very dissimilar texts. To find out if the observed RT difference is caused by the formality inconsistency, I applied Lexile measures (mentioned in Section 5.3.5) in this section and compared the measures with the results of the paired samples t-tests. As none of the pair-wise comparisons in

L2 data was found to be significantly different, the discussion below focuses on the NS data exclusively.

In Pair 1 comparison of Additive MDMs, the paired samples t-tests showed that reading *in addition* in the informal context took longer than reading the same MDM in the formal context. According to Table 5.3, which shows the Lexile measures of all the texts used in Study 1, the Lexile measures for the informal and formal text are 740L and 1230L respectively. These two texts are very different in terms of the Lexile measures and it seems reasonable to assume that the statistical difference found might lie in the difference between the two texts rather than in the use of the marker. However, looking at the comparison in Pair 2 gives a contradictory observation.

Pair 2 compares the informal MDM *what's more* in the same two types of text in Pair 1. If the significance found in Pair 1 is not due to the effect of the formal MDM but to the dissimilarities of the two texts, the comparison in Pair 2 should yield the same result. The comparison in Pair 2, however, showed that the reading time for *what's more* in the formal context ($M = 224.03, SD = 141.93$) was not significantly different from reading the same MDM in the informal context ($M = 186.85, SD = 89.57$), $t(2) = .776, p = .519$. The non-significant finding in Pair 2 suggests that the difference found in Pair 1 is mainly due to the effect of the formality mismatch between the formal MDM and the informal context, rather than to the differences in the two texts.

The same phenomenon was seen in Causal MDMs. The Pair 1 comparison (the formal marker *in consequence* in the 1390L formal passage and in the 840L informal passage) was found to be nearly significant, but not in Pair 2 (informal marker *because*

of *that* in the 1550L formal passage and in the 840L informal passage), $t(2) = 1.107$, $p = .384$.

The result of Pair 1 and Pair 2 with *Concessive* MDMs is a particularly interesting case. According to the Lexile measures, the difficulty or readability levels of the texts are similar, with 1300L for the informal text and 1420L for the formal text (see Table 5.3). The significant differences found in Pair 1 and Pair 2 can thus be attributed to the formality-inconsistency between the markers and the context. The comparison in Pair 1 showed that the RT for the formality-inconsistent text (*despite that* in the informal text) was longer than for the formality-consistent text (*despite that* in the formal context). Although the result of the Pair 2 comparison also showed a significant finding, the RT in the formality-consistent text (*but then again* in the informal context), unexpectedly, was longer than the RT in the formality-inconsistent condition (*but then again* in the formal context). This seemingly contradictory finding will be further discussed in the two-way ANOVA analysis.

A preliminary summary of the paired samples t-tests of the NS and L2 data is that the NS participants noticed the formality inconsistency between the MDMs and the context. L2 participants, on the contrary, did not notice such differences. Specifically, the noticing by NS participants took place primarily in cases of formality mismatch between the formal MDMs and the informal contexts. This phenomenon prevails among the three semantic categories.

5.3.6.5 Two-way ANOVA

The two factors in this analysis are MARKER and REGISTER. The MARKER factor has six levels: formal Additive (*in addition*), informal Additive (*what's more*), formal Causal (*in consequence*), informal Causal (*because of that*), formal Concessive (*despite that*), and informal Concessive (*but then again*). The second factor, REGISTER, has two levels: formal register (FR) and informal register (InfR). The two-way ANOVA was conducted to determine whether or not the six types of marker, when appearing in different registers, would affect the reading time of the manipulated sentences.

NS data

A two-way ANOVA found a main effect of MARKER, with $F(5, 24) = 3.622, p = .014, \eta^2 = 0.24$, indicating that NS participants displayed different response times in reading the six MDMs. There was also a main effect of REGISTER, with $F(1, 24) = 17.644, p < .000, \eta^2 = 0.23$. The difference lay in NS participants' longer reading time for informal passages. In addition, there was an interaction between MARKER and REGISTER, with $F(5, 24) = 3.35, p = .02, \eta^2 = 0.22$. A simple main effects test showed that the response time for formal passages with the various MDMs were the same, $F(2, 24) = 1.380, p = .267$. When reading informal texts, it was found that the NS participants had a longer reading time for Concessive passages, with $F(1, 24) = 5.696, p = .001$ for the formal Concessive marker, and $F(1, 24) = 21.358, p < .000$ for the informal Concessive marker.

L2 data

Another two-way ANOVA run with the L2 data found a main effect of REGISTER, with $F(1, 22) = 7.145, p = .014, \eta^2 = 0.17$. This shows that like NS participants, the L2 participants noticed the formality difference in reading the two types of passage. In addition, the L2 participants also displayed a longer response time in reading informal texts. No main effect of MARKER was found, with $F(5, 22) = 1.017, p = .432, ns$. There was no interaction between MARKER and REGISTER, with $F(5, 22) = 1.800, p = .155, ns$.

5.3.7 Discussion

Both the NS and L2 participants displayed different response times in reading the formal and informal passages. They both spent longer time reading informal texts. When considering text formality and MDMs together, only the NS participants seemed to be aware of the formality-inconsistent conditions, based upon their longer RT in reading formal MDMs in the informal texts. The NS participants' pragmatic awareness is therefore manifested to some extent in the self-paced reading task. In spite of this, their noticing seems to be unidirectional in that the impact of formality mismatch was more pronounced when formal MDMs occurred in the informal passages. When reading informal MDMs in formal contexts, on the other hand, the NS participants did not seem to notice the formality-inconsistency. L2 participants, unlike the NS participants, did not show any noticing of the formality-inconsistency in the task.

An unexpected phenomenon observed from the L2 and NS data is that both groups of participants had an exceptionally longer RT in reading *but then again* in the

informal, Concessive passages, a formality-consistent condition. According to the hypothesis, reading a formality-consistent sentence should take less time than its counterpart, the formality-inconsistent sentence. Items 11 and 12 in Table 5.4 show that both L2 and NS participants spent longer reading formality-consistent cases. Such unusualness might suggest a problem with the material used for this passage. The essence of using corpus-informed material is that the grammar and the wording of the corpus material should remain unchanged. This, however, leads to a *readability* problem with the material. Reading a passage that is full of colloquial usage in a written format might not be what the participants were accustomed to, which explains the longer reading time in this case. Nevertheless, this does highlight the extra care that one needs to take when employing corpus-informed materials.

Another factor that might help to explain the longer reading time for Concessive passages is a particular trait found in Concessive markers. Rudolph (1996) claimed that Concessive markers mark both “the background causality and the rejection of an imaginable conclusion” (p.35). The fact that concession is a more complicated concept to comprehend might incur longer processing time, which in turn explains the longer RT found in the data.

As the method—a reaction time task—used in testing pragmatic awareness has not been previously introduced, Study 1 was an attempt to explore whether the reaction time paradigm is an appropriate methodology to measure pragmatic awareness. The methodology seems to work to some degree, as NS participants did spend more time in responding to the formality-inconsistent conditions. Nonetheless, as this is a new method, the fact that L2 participants did not notice any of the formality

mismatches does not necessarily lead to the conclusion that these L2 participants were unaware of the formality difference. The formality mismatch devised via a reaction time paradigm might be presented in ways too subtle for L2 participants to grasp. Nevertheless, the fact that L2 participants did not show any RT difference in reading the two conditions does suggest an underlying problem in this regard. For example, using the transcription of spoken data (CANCODE) as testing materials poses another problem for L2 participants. According to some L2 participants, reading this informal style of writing is not a familiar reading process, and it could be argued that this unfamiliarity might have influenced their behaviour in pressing the response key. In addition, as the study was designed with the intent of gaining a better understanding of cognitive process and the participants were instructed to read as fast as they could, only three types of MDM were included so as to avoid fatigue effects. Should additional semantic types of MDM be tested, L2 participants might have a different task performance.

5.3.8 Conclusion

By using corpus-informed material coupled with psycholinguistic methodology, I have reported in Study 1 a new approach in investigating L2 learners' pragmatic awareness. Through the self-paced reading task designed in the psycholinguistic paradigm, I explore the possibility of measuring NS and L2 learners' pragmatic awareness by observing their reading behaviour with different MDMs in two types of texts. Despite the non-significant findings in both NS and L2 data, the NS participants did notice the formality-inconsistency in the reading tasks (formal MDMs in informal

texts) and produced longer response times. Further, among the three types of MDM, it seems that the concept of Concessives is the most difficult for all participants to comprehend, as is evidenced by the longest RT in both the formality-consistent and inconsistent conditions. Although I have mentioned that reading colloquial corpus-informed material might pose a problem, it is also evident that understanding the notion of concession imposes a greater cognitive burden, which in turn might prolong readers' reaction time. The longer RT caused by increased mental effort is especially noticeable in the formality-inconsistent Concessive passages.

Despite the evidence and findings reported above, this study is still at its exploratory stage. A reaction time design that directly compares spoken and written data poses one challenge. It is either to change speech into written form or to change written words to an audio input⁶. In order to amend the problems observed, experiment 2 was designed to include five different types of MDM and the two registers compared were both in the written mode: formal writing and informal writing.

5.4 Study 2: A multiple-choice discourse completion task (MDCT)

Study 2 involved a multiple-choice discourse completion task (MDCT), which had ten short passages, each with an MDM missing. Instead of replacing the formal with the informal MDMs, as was the case in Study 1, the ten MDMs were purposely omitted. Participants read the ten passages and selected an appropriate MDM from the provided list for each passage. Five of the ten passages were formal texts and the other

⁶ Including an audio file for the spoken texts would not justify the comparison, as it involves another variable, language modality.

five were informal passages. The ten MDMs (in five pairs), accordingly, included five formal and five informal markers. Each pair represented one semantic category. For example, *as a result* and *because of that* were the two MDMs in the Causal pair.

The hypothesis of this study is that when reading, for example, a formal Causal passage, participants who have noticed the formality of the passage will accommodate to the context and choose the formal Causal MDM, *as a result*. Participants who have only paid attention to the linking marker might simply choose *because of that*, regardless of the contextual formality cues. In addition to testing language users' pragmatic awareness of the register difference, I also explore to what extent this methodology can truly reflect the participants' pragmatic awareness.

5.4.1 Participants

Twenty-nine native speakers of English (NS) and thirty L2 learners participated in the second study. Like Study 1, all of the fifty-nine participants were postgraduates at the University of Nottingham. While the L2 participants in the first study came from various L1 backgrounds, the thirty L2 participants in this study are homogeneous with respect to their L1 background: they are all native speakers of Mandarin Chinese. The L2 participants had been in the UK for nine months on average when taking part in this study. On the basis of their IELTS score, I divided the L2 participants into two groups, the advanced (Adv.) group, with learners who scored 7 to 7.5 in IELTS, and the intermediate (Inter.) group that includes L2 participants whose IELTS score was between 5.5 to 6.5. The NS group comprised twenty-five English, three American and one Canadian participant.

5.4.2 Materials

Traditionally, discourse completion tasks (DCTs) include tightly controlled elicitation techniques and the wording might influence or direct participants' behaviour. The need to include situated authentic discourse data has been pointed out in ILP research (Kasper & Dahl, 1991; Bardovi-Harlig & Hartford, 2005). Following this line of thinking, I used corpus-informed data as the materials.

Because the register comparison in Study 2 was made between formal written and informal written discourse, the materials used were extracted from the written section of the BNC corpus (BNC-W) and the Enronsent2 e-mail corpus. To access BNC-W, I used the open corpora provided by Sketch Engine (Kilgariff et al., 2004). The Enronsent2 corpus, which is available to download online (Klimt & Yang, 2004), is a large collection (1.5 million words) of e-mail messages. Only a few researchers have used this dataset in L2 pragmatic studies (De Felice & Deane, 2009). I included the Enron e-mail dataset in the pursuit of L2 learners' sensitivity to the less formal register. Although the Enronsent2 corpus includes business communication and personal e-mails, I paid extra attention in choosing appropriate passages for this experiment. That is, I purposely avoided formal content that appears in the e-mail corpus.

The ten MDMs examined in this study were the six MDMs looked at in Study 1, two *Enumerative* MDMs from my earlier study (2010) and another two *Appositive* MDMs that I reported in Chapter 3 (see Table 5.8). These ten MDMs were searched in the two corpora in order to find suitable passages to be used in the MDCT.

Table 5.8

The ten MDMs used in the MDCT

Semantic category	Formal MDM	Informal MDM
Appositive	<i>that is</i>	<i>what I'm saying is</i>
Additive	<i>in addition</i>	<i>what's more</i>
Causal	<i>as a result</i>	<i>because of that</i>
Concessive	<i>despite that</i>	<i>but then again</i>
Enumerative	<i>to begin with</i>	<i>first of all</i>

I adopted two criteria when searching for MDMs in the two corpora. First is the position of the MDMs. Biber et al. (1999) examined the positions of linking adverbials (discourse markers in this thesis) in the Longman Spoken and Written English Corpus, which is a collection of forty million words of British and American English, and concluded that the initial position is found to be the most common and “can thus be considered the unmarked position for linking adverbials” (1999, p. 890). Following this finding, only passages with MDMs occurring in the sentence-initial position were treated as potential candidates. The second criterion is that only one MDM occurs in one passage. This is to minimize the possible interpretations of the coherence relation suggested in the passage and to avoid any possible distraction. Two passages were selected for each MDM. As a result, twenty passages were chosen: ten from BNC-W and another ten from the Enron e-mail corpus. All the MDMs occurring in the twenty

passages were replaced with blanks and the order of these passages was randomized.

These twenty passages were presented to five NS for a pilot test.

The purpose of the pilot test was to ensure that only one semantic type of MDM was plausible in the passages included in the final MDCT. The benchmark for deciding which passages to include in the study was whether or not all the five NS chose the same marker for each passage. On the basis of the answers given by the five NS, ten passages were selected for the experiment (see Appendix 4). Note that the ten MDMs were placed at the top of each page (from page 2 to page 5) to facilitate the selection process. To avoid the item order effect, the order of the ten MDMs was different on each page.

Each participant was tested under ten conditions (passages), each being a combination of two factors, one of which was *REGISTERS* (formal or informal style), and the other *SEMANTICS* (Additive, Appositive, Causal, Concessive or Enumerative). The ten passages were arranged in a Latin Square design, counterbalanced in terms of formality and the semantic categories so that the carryover effect would be minimised. Ten versions of the experimental materials were therefore created. Table 5.9 shows the order of the ten passages in List A.

Each passage had a simple True/False question displayed under each item. The True/False questions were designed on the basis of the passage content and thus were irrelevant to the choice of MDMs that participants made. Their purpose was to decrease the possibility that participants might pick up the formality hint in the reading process.

Table 5.9

Item order in the MDCT, List A

Semantic & formality		
Items	category	MDM
1	Formal Appositive	that is
2	Informal Enumerative	first of all
3	Formal Additive	in addition
4	Formal Concessive	despite that
5	Informal Causal	because of that
6	Formal Enumerative	to begin with
7	Informal Concessive	but then again
8	Informal Appositive	what I'm saying is
9	Formal Causal	as a result
10	Informal Additive	what's more

5.4.3 Procedure

Study 2 took place in a lab setting at the University of Nottingham. All participants were given instructions on how to carry out the MDCT, followed by a practice section. In the instructions, participants were told to read each passage carefully and to choose the most appropriate, natural missing phrase for each passage from the ten choices provided. That is, neither *formality* (register) nor *discourse marker* was mentioned in the instructions. Before beginning the task, each participant

was asked if they knew the ten phrases shown in the practice section. All the NS and L2 participants confirmed that they understood the meanings of the ten phrases (MDMs). As it was not a timed task, participants took their time to complete the task. Each participant received a small reward for taking part in the task.

5.4.4 Measures

The collected data were looked at from three angles. The first was whether or not participants had chosen the correct *class* of MDMs, with class referring to the five semantic categories. The second consideration was whether or not participants had identified the correct *register* in each of the passages. The last aspect that I looked at was whether or not participants had identified the *original* MDMs. This refers to the situation where participants identified the exact MDMs used by the original writers of the ten passages.

For example, when a participant has chosen *despite that* for a formal Concessive passage, this participant's choice has satisfied the three considerations: the right class, the right register, and the exact match. This participant shows his noticing of the register difference as well as the semantic category differences. If, on the other hand, this participant has chosen *but then again* for the same formal Concessive passage, this participant has only met the class requirement. In addition, once a participant has identified the exact MDMs that the original writers had used in the texts, this presupposes that this participant has also identified the right register. In other words, even though I have put forward three angles, the exact match and the register factor consider the same aspect. I therefore included two measures in examining the data,

the CLASS measure and the EXACT MATCH measure. All the participants' answers were looked at twice, with the first assessment emphasizing whether the right CLASS of MDMs is chosen, and the second, whether the chosen MDMs are the EXACT MATCH.

After identifying the CLASS measure and the EXACT MATCH measure, I propose that these two measures pave the way for detecting as well as quantifying participants' pragmatic awareness of register difference. When a participant identifies the informal Causal marker *because of that* for a formal Causal passage, this participant has identified the right class but not the right register. In a case like this, this participant shows his *noticing* of the coherence link but he fails in *noticing* the formality differences between the MDM and the context. Considering the CLASS measure, this participant gains a score. Considering the EXACT MATCH measure, this participant does not gain a score. That is, this participant has a higher score in the CLASS measure and a lower score in the EXACT MATCH measure. Examining the differences between the two measures enables us to show this participant's pragmatic awareness of register. This will be further illustrated in Section 5.4.6.

The experiment includes three variables: Register (two levels, formal vs. informal), Class (5 levels), and Group (3 levels, NS, Adv, and Inter L2). Because both NS and L2 participants undertook the same task, it is a repeated measurement. The statistical method used in analysing the data is a three-way ANOVA. The three-way ANOVA was run twice, first with the scores in CLASS, and then with the scores in EXACTMATCH.

5.4.5 Results

I present the three-way ANOVA results in the CLASS measure, and then in the EXACT MATCH measure, followed by a further discussion on how the EXACT MATCH measure leads to the discussion of participants' noticing of register.

5.4.5.1 Three-way ANOVA in CLASS measure

A repeated measures ANOVA was conducted to evaluate the effects of Class, Register, and Group. Table 5.10 shows the mean and the standard deviation of the ten conditions.

A significant main effect was found for Register, $F(1, 55) = 4.543, p = .038, \eta_p^2 = .076$. Pair-wise comparisons showed that informal passages, in general, had higher scores than the formal passages. Another main effect was found for Groups, $F(2, 55) = 10.907, p < .000, \eta_p^2 = .284$, suggesting that the three groups of participants performed quite differently in the task. Pair-wise comparisons showed that while NS participants scored in a similar way to the Adv L2 participants, NS participants had a higher task score than the Inter L2 group, at $p < .000$. The task performance between the Adv L2 learners and the Inter L2 learners was found to be nearly significant, at $p = .063$.

Another main effect was found in the variable Class. Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2 = 19.172, p < .05$). Degrees of freedom were therefore corrected by using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .842$).

The effect found in Class suggest that the five semantic categories had very different scores: $F(3.369, 185.304) = 4.155, p = .005, \eta_p^2 = .070$.

Table 5.10

Mean and standard deviation of the ten conditions in the CLASS measure, F: formal, Inf:

informal

	Participants	Mean	Std. Deviation	N
F Additive	NS	.9310	.25788	29
	Adv L2	.8182	.40452	11
	Inter L2	.7222	.46089	18
Inf Additive	NS	.6897	.47082	29
	Adv L2	.5455	.52223	11
	Inter L2	.4444	.51131	18
F Appositive	NS	.3793	.49380	29
	Adv L2	.8182	.40452	11
	Inter L2	.5556	.51131	18
Inf Appositive	NS	.7931	.41225	29
	Adv L2	.8182	.40452	11
	Inter L2	.7778	.42779	18
F Causal	NS	.7931	.41225	29
	Adv L2	.7273	.46710	11
	Inter L2	.7222	.46089	18
Inf Causal	NS	.9310	.25788	29
	Adv L2	.6364	.50452	11
	Inter L2	.4444	.51131	18
F Concessive	NS	.8966	.30993	29
	Adv L2	.7273	.46710	11
	Inter L2	.2222	.42779	18
Inf Concessive	NS	.5172	.50855	29
	Adv L2	.2727	.46710	11
	Inter L2	.5000	.51450	18
F Enumerative	NS	.2414	.43549	29
	Adv L2	.2727	.46710	11
	Inter L2	.0556	.23570	18
Inf Enumerative	NS	1.0000	.00000	29
	Adv L2	.9091	.30151	11
	Inter L2	.9444	.23570	18

The results also showed that an interaction was found between Class and Group: $F(8, 220) = 2.496, p = .013, \eta_p^2 = .083$. A simple main effects test indicated that the differences lay mainly in the NS group and the Inter L2 group's scores in reading the Causal passages ($p = .008, \eta_p^2 = .162$), and in reading the Concessive passages ($p = .001, \eta_p^2 = .211$). Another interaction was found when taking into account the three variables, Class, Register, and Group: $F(8, 220) = 3.075, p = .003, \eta_p^2 = .101$. A simple main effects test showed that NS participants gained a better score than the Inter L2 group when reading the informal Causal passage, $p = .000, \eta_p^2 = .236$, and when reading the formal Concessive passage, $p = .000, \eta_p^2 = .391$. Another difference was found when comparing the Adv L2 group with the NS participants. The Adv L2 participants scored better than the NS participants in reading the formal Appositive passages: $p = .040, \eta_p^2 = .109$. When compared the two L2 groups, the Adv L2 participants outperformed the Inter L2 groups in reading the formal Concessive passage: $p = .003, \eta_p^2 = .391$.

5.4.5.2 Three-way ANOVA in EXACT MATCH measure

Another repeated measure ANOVA was conducted to evaluate the effects of Class, Register, and Group within the EXACT MATCH measure. The descriptive statistics are shown in Table 5.11.

Table 5.11

Mean and standard deviation of the ten conditions in EXACT MATCH measure, F: formal,

Inf: informal

	Participants	Mean	Std. Deviation	N
F Additive	NS	.7931	.41225	29
	Adv L2	.6364	.50452	11
	Inter L2	.5000	.51450	18
Inf Additive	NS	.4828	.50855	29
	Adv L2	.4545	.52223	11
	Inter L2	.3333	.48507	18
F Appositive	NS	.3448	.48373	29
	Adv L2	.3636	.50452	11
	Inter L2	.5000	.51450	18
Inf Appositive	NS	.7931	.41225	29
	Adv L2	.6364	.50452	11
	Inter L2	.6111	.50163	18
F Causal	NS	.4828	.50855	29
	Adv L2	.0909	.30151	11
	Inter L2	.3889	.50163	18
Inf Causal	NS	.7241	.45486	29
	Adv L2	.5455	.52223	11
	Inter L2	.2778	.46089	18
F Concessive	NS	.8966	.30993	29
	Adv L2	.6364	.50452	11
	Inter L2	.2222	.42779	18
Inf Concessive	NS	.5172	.50855	29
	Adv L2	.0909	.30151	11
	Inter L2	.3889	.50163	18
F Enumerative	NS	.1724	.41225	29
	Adv L2	.1818	.40452	11
	Inter L2	.0556	.42779	18
Inf Enumerative	NS	.7931	.43549	29
	Adv L2	.8182	.40452	11
	Inter L2	.7778	.23570	18

The three-way ANOVA showed that the Class effect was not significant: $F(4, 220) = 1.244, p = .292, \eta_p^2 = .022$. A main effect was found with the Register variable: $F(1, 55) = 7.887, p = .007, \eta_p^2 = .125$. Scores in the informal register were greater than those in the formal register. The effect of the Group variable was found to be significant: $F(2, 55) = 14.631, p < .000, \eta_p^2 = .347$. Pairwise comparisons showed that NS participants had higher scores in the task than did the Adv L2 participants ($p = .004$) and the Inter L2 participants ($p < .000$). No significant difference was observed between the two learner groups ($p = 1.00, ns$).

The three variables, Class, Register, and Group, were also found to have an interaction effect: $F(8, 220) = 2.979, p = .003, \eta_p^2 = .098$. A simple main effects test showed that the effects mainly lay in the informal Causal and the formal and informal Concessive passages. NS participants had a higher score than the Inter L2 participants in identifying the missing MDM in the informal Causal passage: $F(2, 55) = 5.017, p = .008, \eta_p^2 = .154$. When finding the Concessive MDM in the formal register, NS participants outperformed the Inter L2 participants, with $F(2, 55) = 16.640, p < .000, \eta_p^2 = .377$. Likewise, the Adv L2 participants also had a higher identification rate than the Inter L2 participants in this case, with $F(2, 55) = 10.640, p = .022, \eta_p^2 = .377$. In the informal register, NS participants showed better identification of the informal Concessive MDM than the Adv L2 participants, with $F(2, 55) = 3.211, p < .043, \eta_p^2 = .105$.

5.4.5.3 Discussion of the statistical analysis

Running the three-way ANOVA in the CLASS and the EXACT MATCH measure, I found that both Register and Group had an effect on the task results. In general, passages in the informal register had a higher score than did passages in the formal register. NS and Adv L2 participants performed equally well in the CLASS measure. At the same time, the scores of the Adv and the Inter L2 participants were marginally significant ($p = .06$). When the data was analysed in the EXACT MATCH measure, NS showed better task performance than both the Adv and the Inter L2 participants. No difference was observed between the Adv and the Inter L2 participants. Interpreting this with regard to the register consideration, it seems that while the Inter L2 participants noticed neither the register nor the semantic differences, which resulted in low scores in both the EXACT MATCH and the CLASS measure, the Adv L2 participants were able to identify the right class of the missing MDMs, yet less able to discern the register difference. The observed difference between the two L2 participant groups suggests that the level of English proficiency affects L2 participants' noticing level.

Both of the ANOVA analyses showed an interaction effect among the three variables, suggesting that the three groups of participants identified the missing MDMs differently in the formal and the informal register. Of the five semantic categories, the Concessive seemed to be the most problematic for the two learner groups. Specifically, the low scores of the Inter L2 participants in identifying the Concessive MDMs in the formal and the informal register led to the significant differences observed when compared with the NS and the Adv L2 participants. The Adv L2 group, on the other

hand, did not show a problem in identifying the formal Concessive MDMs. Rather, they showed a low identification score in reading the informal Concessive passage, which resulted in the significant difference found between the Adv L2 and the NS participants.

5.4.5.4 Implication of the mean scores in the CLASS measure

After discussing the results of the statistical analysis, in this section I consider the implication of the CLASS measure and propose that participants' mean scores in this measure suggest the ease or difficulty of the semantic linking type. Table 5.12 shows participants' total mean scores for the five semantic types in the two registers.

Table 5.12

Total mean scores for the five semantic types

F + Inf			
Register	NS	Adv L2	Inter L2
Additive	1.62	1.36	1.17
Appositive	1.17	1.64	1.33
Causal	1.72	1.36	1.17
Concessive	1.41	1.00	0.72
Enumerative	1.24	1.18	1.00
Sum	7.16	6.54	5.39

The sum column shows that NS participants had the highest mean scores, followed by the Adv L2 and Inter L2 participants, indicating that English proficiency affects the identification of semantic ties. Ranking the mean scores of the five semantic types in the three groups, I discovered that the L2 participant groups had an identical ranking order. They both had the highest mean scores for Appositive MDMs, followed by Causal, Additive, Enumerative, and Concessive. The NS participants, on the other hand, scored the highest in Causal MDMs, followed by Additive, Concessive, Enumerative, and Appositive. The order discrepancies and the fact that both the Adv and the Inter L2 participants displayed a homogeneous character in the ranking order suggest that L2 learners' order of acquisition of the semantic ties is different from that of NS. Identifying the Concessive link seems to be the most challenging for the L2 participants, whereas identifying the Appositive link appears to be the easiest. The fact that the highest mean scores were achieved in identifying the Appositive link, however, is not consistent with the CLC corpus finding, in which the Appositive MDMs were used in a very limited manner. Such a discrepancy between the identification of MDMs (a cognitive process) and the employment of MDMs (a behavioural process) will be further addressed in Chapter 6.

5.4.5.5 Difference between the CLASS measure and the EXACT MATCH measure

In this section, I show how the measures of CLASS and EXACT MATCH lead to the discovery of participants' pragmatic awareness of the formality differences in the two registers. To illustrate how this is done, I will start by looking at the participants' two mean scores in the informal Causal condition.

When reading the informal Causal passage, participants gain one score in both the CLASS and the EXACT MATCH measures if they identify the informal Causal MDM *because of that*. When participants do not notice the informal register of the passage but still identify the Causal link by choosing the formal Causal MDM *as a result* for this condition, they gain one score under the CLASS measure. In other words, the CLASS measure includes cases where participants identify the formal and the informal MDM that fall into the same semantic category.

The mean scores of this condition in the EXACT MATCH and the CLASS measure as well as the two-measure differences are shown in Table 5.13.

Table 5.13
Mean score of the two measures and the difference in the informal Causal condition

Measure	NS	Adv L2	Inter L2
EXACT MATCH	0.72	0.55	0.28
CLASS	0.93	0.64	0.44
Difference	-0.21	-0.09	-0.16

The Difference row in Table 5.13 indicates the extent to which the participants show a lack of pragmatic awareness of register difference. A greater mean score difference suggests a high mean score in the Class measure but a low mean score in the EXACT MATCH measure. A lower mean score in the EXACT MATCH measure means that participants do not notice the formality or the register difference. That is, a greater gap between the two measures would naturally suggest a greater lack of

pragmatic awareness. Although the NS participants showed the greatest difference between the two measures (-0.21), they scored the highest in both the CLASS and the EXACT MATCH measure. Despite a smaller mean score difference (-0.09) in the Adv L2 participants, they had lower mean scores in the two measures than did the NS participants. Inter L2 participants, on the other hand, scored the lowest in both of the measures. The low mean scores in the two measures, together with the mean score difference of -0.1666, suggest that the Inter L2 participants, when compared with the other two groups of participants, do not yet have a firm grip on the formality difference. Using the informal Causal condition as an example, I show that the difference between the two measures leads to the discovery of participants' lack of pragmatic awareness, and the interpretation of the two mean score differences will have to take into account the CLASS and EXACT MATCH mean scores.

Following the same procedure, Table 5.14 shows the results of the subtraction of the two measures in the formal conditions and Table 5.15 shows the differences in the informal conditions.

All of the twenty-nine NS participants chose the Concessive markers in accordance with the register formality, regardless of whether this involved choosing *but then again* for the informal passage or *despite that* for the formal passage, shown by the zero difference in Tables 5.14 and 5.15.

Table 5.14

Mean score differences of the two measures in the five formal conditions

Condition	Formal Register					
	(EXACT MATCH - CLASS)					
	NS		Adv L2		Inter L2	
Additive	(0.79-0.93)	-0.14	(0.64-0.82)	-0.18	(0.50-0.72)	-0.22
Appositive	(0.34-0.38)	-0.03	(0.36-0.82)	-0.45	(0.50-0.56)	-0.06
Causal	(0.48-0.79)	-0.31	(0.09-0.73)	-0.64	(0.39-0.72)	-0.33
Concessive	(0.90-0.90)	0.00	(0.64-0.73)	-0.09	(0.22-0.22)	0.00
Enumerative	(0.17-0.24)	-0.07	(0.18-0.27)	-0.09	(0.06-0.06)	-0.00
Sum		-0.55		-1.45		-0.61

The sum columns in the two tables show the extent to which the participants overlooked the formality or register difference. NS participants had a smaller sum in the formal register (-0.55) than in the informal register (-0.63). This suggests not only that the NS participants noticed the micro-level of semantic differences within various MDMs, but also that they noticed the macro-level of the formality differences in the context, and made their choice of markers accordingly. The only condition in the formal register that had a greater mean score difference was the formal Causal passage, and this was due to the fact that nine NS participants selected the informal Causal MDM *because of that*.

Table 5.15

Mean score differences of the two measures in the five informal conditions

Informal Register						
(EXACT MATCH - CLASS)						
Condition	NS		Adv L2		Inter L2	
Additive	(0.48-0.69)	-0.21	(0.46-0.55)	-0.09	(0.33-0.44)	-0.11
Appositive	(0.79-0.79)	0.00	(0.64-0.82)	-0.18	(0.61-0.78)	-0.17
Causal	(0.72-0.93)	-0.21	(0.55-0.64)	-0.09	(0.28-0.44)	-0.16
Concessive	(0.52-0.52)	0.00	(0.09-0.27)	-0.18	(0.39-0.50)	-0.11
Enumerative	(0.79-1.00)	-0.21	(0.82-0.91)	-0.09	(0.78-0.94)	-0.16
Sum		-0.63		-0.64		-0.71

The Adv L2 participants had very different sums of mean score differences in the two registers, with -1.45 in the formal register and -0.64 in the informal register. The greater sum in the formal register indicates that when reading formal passages, the Adv L2 participants did not notice the formality of the passages. This is particularly salient in the formal Causal passage, where 88 per cent (0.64/0.73) of the answers given for this condition were the informal Causal MDM *because of that*. Another condition that shows a similar pattern is the formal Appositive passage. In this case, more than half (55%) of the MDMs that the Adv L2 participants chose were the informal Appositive MDM *what I'm saying is*.

Looking at the Inter L2 mean score differences in Tables 5.14 and 5.15, we see that the Inter L2 participants had a pattern that is contrary to the Adv L2 participants.

The Inter L2 participants had a higher sum of mean score differences in the informal register, suggesting that the Inter L2 group tended to choose formal MDMs when reading the informal passages. In addition, I observed a repeated pattern in the formal Causal condition. Like the trend observed among the NS and the Adv L2 participants, nearly half (46%) of the Causal MDMs that the Inter L2 participants identified for this formal Causal passage were the informal Causal MDM *because of that*. That is, there is a general tendency for the three groups to choose the informal MDM for the formal Causal passage. Such tendency indicates that this formal Causal passage itself might be problematic. I elaborate more on this methodological issue in the following section.

5.4.6 Discussion

With the manipulation of formality difference between discourse segments and the MDMs that occur in them, I have shown a new method for detecting L2 learners' pragmatic awareness or sensitivity in two registers of written discourse, the formal and the informal. It was found that the NS participants accommodated to the formal passages and chose the formal MDMs for these passages accordingly. The L2 learners, on the contrary, tended to choose informal MDMs for the formal passages, and this was particularly salient in the Adv L2 participants. The results suggest that L2 participants have formed a global impression of passage formality, but they have not yet grasped the formality that lies in the phrasal or sentence level.

A reasonable prediction followed by NS participants' noticing of the formality in the formal register is that they would also have a keen awareness in selecting appropriate MDMs in the informal register. This is because formal language is usually

more difficult to comprehend than informal language. However, this was not the case. Such a discrepancy can be explained via the mental effort incurred during the reading tasks. Britton et al. (1982) reported that “text versions with simplified vocabulary and syntax (but equivalent content) required less cognitive capacity to process than standard versions” (p. 51). Reading formal academic passages, as a result, imposes a heavier mental load. When undertaking a reading task in which both formal and informal texts are included, it is possible that participants expend a great deal of mental effort in reading the formal passages. Such forced effort might therefore have a carry-over effect when participants read informal passages.

This mental load effect was also seen in L2 participants but at a different level. Reading texts in a non-native language put the L2 participants at a relative disadvantage in relation to the NS participants. Much of the mental effort of the L2 participants might be spent in comprehension and in finding the missing phrase (MDM). The macro-level noticing of the formality difference, therefore, might have been suppressed. This is particularly evident in the Adv L2 participants' task performance in the formal register.

The method of utilizing corpus-informed material in the MDCT design has proved to be a plausible approach; however, some amendment and improvement is needed. The ten short texts, although selected after conducting a pilot test, were presented to the participants as they were when selected from the corpora. The grammar and the sentences remained virtually intact based on the underlying assumption that the corpus-informed material should be kept original. Some of the L2 participants reported that at times they encountered very difficult words during the reading

process. At the same time, some reported that they were not accustomed to reading colloquial English. Both of these factors might affect their judgment in selecting the appropriate markers. This flaw of employing authentic corpus material should be dealt with when replicating this method. In addition, the very low mean scores across the three participant groups in the formal Enumerative condition and the greater mean score differences in the formal Causal condition all suggest that the selection of the material and the pilot test have to be carried out with more care. Focusing exclusively on Chinese learners of English has appeared to be valuable in finding a generalized pattern of L2 learners. A good example is that the Adv and Inter L2 participants showed an identical ranking order in the mean scores they obtained in the five semantic types of passages. Both the L2 participant groups had the highest mean scores for the Appositive MDMs, followed by Causal, Additive, Enumerative, and Concessive. Despite this finding, I did not pursue further the influence that L1 had on their choice of MDMs. Whether or not L1 has an influence on L2 participants' choice of MDMs should be addressed in future studies.

5.5 Putting Studies 1 and 2 together

I designed Studies 1 and 2 to investigate L2 participants' sensitivity in two registers: formal and informal discourse. In Study 1, a self-paced reading time experiment was introduced to see how NS and L2 learners read formality-inconsistent sentences. The rationale of the design is that reading formality-inconsistent sentences would produce a longer reaction time. NS participants did display a longer response time to formality-inconsistent conditions; the prolonged reading time was shown

particularly in reading informal passages with formal MDMs. Like NS participants, the L2 participants noticed the formality differences in the two different types of passage. However, they did not perceive the subtle formality differences between MDMs and the sentences where these markers appeared. It seems that for the L2 participants in Study 1, a global impression of the formality was created but not a local impression where the formality lies in the phrasal or sentence level.

In Study 2, I designed a multiple-choice discourse completion task (MDCT) in which NS and L2 participants read ten passages and chose one natural phrase to insert into each passage. NS participants accommodated to the formal passages, tending to choose the formal MDMs for the formal passages accordingly. While the Adv L2 participants identified more informal MDMs for the formal passages, the Inter L2 participants scored low in both the formal and informal registers. In other words, we observe that English level plays a role in how much attention is paid to the formality difference. In terms of the five semantic types of MDM tested, while NS participants showed the highest mean score in identifying the Causal link and the lowest in the Appositive, both the Adv and Inter L2 participants found identifying the Appositive link the easiest but the Concessive the most difficult.

One difference observed between Studies 1 and 2 is the global impression of formality. The results of the two-way ANOVA run in Study 1 show that the L2 participants noticed two types of discourse in the RT task: formal and informal. The results of the three-way ANOVA in Study 2 do not indicate such noticing. Two explanations are offered for this difference. The two registers compared in Study 1 were formal written and informal spoken discourse. Both of the registers compared in

Study 2 were in written mode, one formal and the other informal. The fact that the informal discourse in Study 2 was in written mode, which might not be as colloquial as informal spoken discourse, might explain why L2 participants in Study 2 failed to notice the formality difference between the two registers. Another possible reason is nature of the two studies. Study 1 was a purely reading comprehension task whereas Study 2 involved not only comprehension but also an action: choosing an MDM for each passage. L2 participants might have simply focused their attention on trying to understand the passage, and on the task of selecting a missing MDM.

The methodologies used in Studies 1 and 2 are a new attempt in testing language users' pragmatic awareness of register difference. Although both of the studies have yielded interesting results, the new methods, combined with corpus-informed materials, have shown that some modification is needed. Low-frequency words or sentences that are written in colloquial style have been proven to affect L2 participants' task performance. Employing a reaction time task requires a sophisticated design of the testing material, and yet I used corpus-informed material in this pursuit. This combination runs the risk of compromised quality control, as the passages used in the study were produced by various writers. Although the reaction time task has been shown to demonstrate NS participants' pragmatic awareness to some extent, it does not achieve the same effect when readers are L2 learners. The MDCT design, like Study 1, also suffers from employing unedited corpus material. Although not including filler items might be one methodological flaw, the un-timed reading nature of the task suggests that a longer time on task would increase the fatigue effect on participants.

The experimental findings as well as the discussion on methodologies have provided evidence to answer the fifth research question raised in Chapter 1.

RQ5: To what extent can we measure and quantify learners' pragmatic awareness in registers via a reaction-time paradigm and an MDCT device?

While both the reaction-time design and the MDCT task successfully reflected NS participants' pragmatic awareness of register difference, only the MDCT task showed L2 learners' awareness. Although the rationale behind the reaction-time design seems intuitively probable, the fact that such a design did not reflect L2 participants' insensitivity to register points to two directions for possible improvement. Because the reaction-time paradigm showed NS participants' awareness to some extent, it is possible that involving L2 participants with a higher English proficiency level in the same design might lead to a different result. Another direction would be to refine the approach by, for example, having the participants respond on a word-by-word basis, rather than on a sentence-by-sentence basis. That is, each slide would show a word instead of a sentence. The RT would be recorded as milliseconds per word.

5.6. Summary and conclusion

I have presented two studies, exploring L2 pragmatic awareness in the use of MDMs in two registers. The reaction time paradigm used in Study 1 compared academic written discourse with conversational spoken data. The second study, which was presented to participants in an MDCT format, compared formal written discourse

with informal writing. NS in the two experiments tended to notice the formality-inconsistent cases, no matter whether the inconsistency was presented via a reaction time task or an MDCT format. The L2 participants, on the contrary, only noticed the formality of the larger passages, failing to notice the formality differences embedded at the sentence level.

Despite the methodological issues that I have mentioned, the two studies make a contribution in seeking psycholinguistic evidence with regard to L2 learners' awareness of register difference. Unlike conventional experiments that include material designed to fit into the research purpose, I took a new approach by including authentic corpus material in both studies. Despite the limitations mentioned earlier, the two study designs, coupled with corpus methodology, have proven to be a plausible approach in unfolding the potential that a combined methodology has in exploring L2 pragmatic competence.

In the next chapter, I will compare the findings in this chapter with the learner corpus evidence obtained in Chapter 4 and discuss what insights we can gain into L2 pragmatic awareness.

Chapter 6 - Putting corpus evidence and experiments together

I have taken two different approaches in examining L2 learners' use of MDMs in the formal and informal registers. The first approach was a corpus-based method in which I analysed the L2 learner data obtained from the Cambridge Learner Corpus (CLC) and compared it with NS data. The corpus evidence suggests that learners' problem in using MDMs in correct registers cannot be described simply by frequencies of MDMs. Learners' problem of using MDMs with sensitivity to registers has to be examined according to the semantic categories of the MDMs. The second approach taken was an experimental one in which I designed two studies—a reaction time task and a multiple-choice discourse completion task (MDCT)—to test L2 learners' pragmatic awareness with regard to their use of MDMs in two registers.

After presenting the learner corpus data (Chapter 4) as well as the experimental results from the two studies (Chapter 5), in this chapter I aim to make a thorough evaluation and comparison of the two general approaches taken. The corpus-based approach provides authentic L2 language data for examining L2 learners' use of various MDMs, whereas the experimental approach offers a valuable means to work empirically on the subject—L2 pragmatic awareness in using MDMs in two different registers. Despite the differing rationale between the two approaches, the relationship between them will contribute to a deeper understanding of the issue in discussion. There are two goals in this comparison process. The first goal is to see whether or not L2 learners' usage pattern of MDMs found in the learner corpus correlates with the L2 participants' task performances in the two studies. In addition, the comparison of the

two types of data serves a useful purpose in evaluating whether the methodologies adopted in the experiments are adequate to measure L2 pragmatic awareness.

Bringing together the empirical evidence and the corpus data, I aim to evaluate the innovative experimental methods used, and to provide a solid basis from which to put forward my understanding of learners' pragmatic awareness of register difference.

This chapter will be structured as follows. First, the results of the two studies will be compared with the learner corpus findings. On the basis of these comparisons, I will evaluate the methodologies used in the two studies. As the two studies employed two different methods, the evaluation will focus on each method's plausibility and applicability.

6.1 Study 1 and the corpus evidence

6.1.1 A quick recap of Study 1

Study 1 involved a reaction time (RT) task in which participants were asked to read eighteen short passages in front of a computer monitor. Six of the passages were written in formal style, six in colloquial English, and another six passages were composed with a neutral tone. Each passage included a target sentence, which was devised to create either a formality-consistent or a formality-inconsistent condition. The formality-consistent condition refers to the situation where the formality level of the target sentence and the embedded MDM are at the same level. The formality-inconsistent condition, on the other hand, includes sentences where the formality of the MDM and that of the context do not match. Examples of these two conditions are given in sentences (1) and (2).

(1) We've got good parks. *But then again*, the kids won't play in the parks.

(2) We've got good parks. *In spite of that*, the kids won't play in the parks.

These two sentences are identical, with the exception that the linking Concessive MDMs are different from one another. As the context suggests a very informal style of writing, sentence (1), which is composed with the informal Concessive MDM, is an example of the formality-consistent condition, whereas sentence (2), written with the formal Concessive MDM, illustrates the formality-inconsistent condition.

Passages in Study 1 were designed to be shown on slides, with one or two sentences displayed per slide. In order to read the next slide, participants pressed a response entry key. In other words, the total reading time for each sentence was recorded by calculating the duration between the two presses. The RT in milliseconds (ms.) used in data analysis was calculated by taking the total reaction time of the target sentence and dividing it by the number of syllables included in the target sentence. The hypothesis of this design is that participants with fully developed pragmatic awareness will find the formality-inconsistent sentence odd and would hesitate to press the entry key.

6.1.2 Interpreting the meaning of long and short RT

Participants' response time in the experimental conditions were analysed via paired samples T-tests and two-way ANOVA in Chapter 5. Specifically, the focus in

Chapter 5 was to compare L2 participants' RT with that of NS participants. In this section, the RT of L2 participants will be compared with the CLC corpus data.

The time that participants take to press the entry key has its implications for the understanding of learners' pragmatic awareness, as well as their acquisition of different MDMs. As this study includes both the formality-consistent and formality-inconsistent conditions, the inferences drawn from the RTs in the two situations are rather different.

In conditions where the passages have formality-consistent sentences, a longer RT would suggest that reading the passage as well as the sentence where the MDM occurs is more difficult. A shorter RT, on the contrary, suggests that participants find the composition and the linking coherence easier to comprehend. In other words, the RT of formality-consistent conditions serve as indices to show how easy or difficult the target sentences and the cohesive ties are.

The RT in formality-inconsistent conditions has a different implication. A longer RT found for formality-inconsistent conditions has two interpretations. Like the long RT in the formality-consistent case, a longer RT in formality-inconsistent sentences might also suggest that these sentences are difficult to understand. The embedded formality mismatch might or might not have any influence on the RT. The second interpretation is that it is the embedded formality mismatch that leads to a longer RT. In the latter case, this shows that participants display greater pragmatic awareness by noticing the formality difference. A shorter RT, on the other hand, suggests that participants do not perceive the embedded formality mismatch, and therefore, their pragmatic awareness is low.

The inferences drawn from the RT in the two conditions will be compared with the frequencies and types of the MDMs retrieved in the CLC corpus.

6.1.3 Formality-consistent conditions

The six formality-consistent conditions included three formal and three informal passages. For the sake of convenience, I show again the RT of the six conditions in Table 6.1. MDMs that belong to the same semantic category are paired together. To provide a common ground for comparisons of the six texts, I include in Table 6.1 the Lexile measures reported in Chapter 5. As Study 1 only includes three semantic categories, I will only refer to the corpus data that are related to these three categories.

Although the RT of the NS participants is not shown in this section (see Chapter 5, Section 5.3.6.1), both the NS and the L2 participants showed a longer RT when reading the informal MDMs in the informal context. This might be due to the participants' unfamiliarity with reading spoken language in the written form. Nevertheless, by combining the RT of both the formal and the informal condition, we have the total RT of the formality-consistent conditions in each of the three semantic categories. As discussed earlier, a longer RT for a formality-consistent condition indicates a more difficult semantic type of MDM, whereas a shorter RT implies an easier type of MDM. The sum of Mean RT in Table 6.1 indicates that L2 participants have the longest RT in reading the Additive passages, followed by the RT of the Concessive and the Causal passages. This seems to suggest that if we rank the three types of passage from easy to difficult by mean RT, we have the following order: Causal, Concessive, and Additive.

Table 6.1

RT of the six formality-consistent conditions. F: formal, Inf: informal, C: context

Conditions	L2 mean RT (ms.)	Lexile measures
1. FAdd-FC	181.937	1230L
2. InfAdd-InfC	281.023	740L
Sum	462.960	1970L
3. FCau-FC	198.307	1390L
4. InfCau-InfC	226.637	840L
Sum	424.944	2230L
5. FConce-FC	181.063	1420L
6. InfConce- InfC	270.127	1300L
Sum	451.190	2720L

The sum of Lexile measures show that while Concessive passages have the highest Lexile measures among the three, Additive passages have the lowest. A high Lexile measure indicates a more difficult text to read. Ranking the three types of passage by their Lexile measures from low to high gives rise to the order Additive, Causal, and Concessive. This order also means that the Additive texts are the easiest to read whereas Concessive passages are the most difficult. Such an order, however, is different from the easy-to-difficult order found above from L2 participants' mean RT. The order differences show that a discrepancy exists between the L2 participants' response time and the Lexile measures of the Additive passages.

The corpus data in Chapter 4 (Section 4.4.2) shows that the CLC learners in general have grasped the use of formal Additive MDMs in the formal register. Using informal Additive MDMs in the informal register, however, is found to be more problematic. The long RT in condition 2 (Table 6.1), where L2 participants read the informal sentence in which the informal Additive MDM is embedded, reflects such a difficulty. In fact, the RT in this condition, 281.023, is the longest among the three informal passages. Another piece of evidence that supports the argument that using informal Additive MDMs in the informal register is problematic for L2 learners comes from the Lexile measure. According to the Lexile Analyser, the passage in Condition 2 has the lowest measure, indicating that it is the easiest to read. The L2 participants' task performance in this regard, however, shows the contrary. In other words, the informal Additive passage is the reason for L2 participants' longest RT in reading Additive passages.

In terms of the Concessive passages, the formality distribution of the CLC corpus data shows that L2 learners used formal and informal Concessive MDMs quite infrequently. In addition, more informal Concessive MDMs were found to appear in the formal writing, whereas more formal MDMs were found in the informal writing. The long Sum RT in reading Concessive passages in Table 6.1 is evidence of L2 learners' problems in this respect. This is also supported by the higher Lexile measures of the Concessive passages. The CLC corpus data shows that L2 learners did not show particular problems in using Causal MDMs in either of the two registers. This might suggest that acquiring Causal MDMs is comparatively easier than the other two types of MDM. The Sum RT in condition 3 and 4 strengthens this inference.

6.1.4 Formality-inconsistent conditions

Contrary to Section 6.1.3, in which the RT is defined to show how easy or difficult the cohesive ties are, the RT in formality-inconsistent condition indicates to what extent the participants' pragmatic awareness is raised. Following the presentation in Table 6.1, I summarize the RT of formality-inconsistent conditions in Table 6.2. The Lexile measures of these six passages are also provided. Passages that include the same semantic type of MDM are paired. What is crucial here is the relationship between the RT and the level of pragmatic awareness. The Sum rows are therefore not included in Table 6.2.

Table 6.2

RT of the six formality-inconsistent conditions. F: formal, Inf: informal, C: context

Conditions	L2 mean RT (ms.)	Lexile measures
1. InfAdd-FC	182.563	1220L
2. FAdd-InfC	160.810	740L
3. InfCau-FC	206.343	1550L
4. FCau-InfC	214.045	840L
5. InfConce-FC	139.780	1420L
6. FConce- InfC	211.443	1300L

Other than condition 2, which represents the informal passage with the formal Additive MDM, the L2 participants displayed a longer RT in reading the informal

passages. Ranking the RT of the informal passages by duration, the L2 participants had the longest RT in reading the Causal passage (condition 4), followed by the Concessive passage (condition 6), and the Additive passage (condition 2). The Lexile analyser assigns 840L to the informal Causal passage, 1300L to the informal Concessive passage, and 740L to the informal Additive passage. One might argue that the readability of these texts might affect L2 participants' reading speed. If this is indeed the case, the Concessive passage should have the longest RT and the Additive the shortest; yet, because of the formality mismatch in these three informal passages, the order of RT length from long to short is: Causal, Concessive, and Additive. This means that the formality mismatch has affected participants' reading speed. As I have argued that a longer RT in formality-inconsistent conditions suggests that more pragmatic awareness is involved, the fact that the L2 participants' response time was the longest in reading the informal Causal passage would imply that the formality mismatch in this passage was the easiest to spot; likewise, the shortest RT found in the informal Additive passage indicates that the formality-inconsistency in this case was the least noticed.

Comparing the above finding with the results reported in Section 6.1.3 gives us a clear indication of how aware the L2 participants were of the formality mismatch embedded in the passages they read. Section 6.1.3 shows that the L2 participants spent the shortest time in reading the Causal passage but the longest on the Additive passage, suggesting that the Causal passage is easier to understand than the Concessive and the Additive passage. In other words, when the three types of passage include a formality-inconsistent sentence, such a formality mismatch should be the

most noticeable in the Causal passage but the least in the Additive. The RT length order from long to short in the formality-inconsistent cases (Causal, Concessive, Additive) supports the RT order observed in the formality-consistent conditions.

Ranking the formal passages—which are conditions 1, 3, and 5—by the length of the RT from the longest to the shortest gives rise to the order Causal, Additive, and Concessive. This order suggests that when reading formal passages in which a formality mismatch is devised via the inclusion of informal MDMs, L2 participants' attention to such mismatch is most noticeable in the passage with the informal Causal MDM. The passage with the informal Concessive MDM, on the other hand, only receives scant attention from the L2 participants. Ranking the Lexile measures of these three formal passages from high to low gives us the order: Causal (1550L), Concessive (1300L), and Additive (1220L), suggesting that the Causal passage is more difficult to read than the other two passages. Following this line of reasoning, the formality-mismatch in the formal Causal passage might not be easily perceived. This means that it cannot be discerned whether the long RT in this passage is due to the high Lexile measure or to the noticing of the formality mismatch leading participants to hesitate to press the response key. Nevertheless, by combining the RT observed in the informal Causal passage, it seems that Causal passages, in general, are easier to understand, which gives L2 participants room to expend their mental effort in noticing the formality-inconsistent device.

The amount of pragmatic awareness raised, shown by the length of the RT, corresponds to the CLC corpus evidence. The corpus data shows that while CLC learners' use of Concessive MDMs is found to be problematic, their use of Causal MDMs

does not show particular problems. Even though only three semantic types of MDM are included in Study 1, the experiment results are consistent with the learner corpus evidence.

6.1.5 The RT differences between the two types of conditions

I have pointed out in Chapter 5 (Section 5.3.6.3) that by comparing the same passage in two conditions, a formality-consistent condition and a formality-inconsistent condition, the RT difference would best quantify the amount of pragmatic awareness raised in the experiment. Take NS participants' RT in reading the formal Additive passages for example. The RT of the formality-consistent condition, i.e., the passage where the formal Additive MDM was included, was 116.200ms. When the same formal passage occurred with an informal Additive MDM, NS participants had a longer RT, 224.027ms. Despite the fact that it was the same passage that was read, the different Additive MDMs resulted in a difference in RT of 107.827ms. This difference in turn suggests the pragmatic awareness that is incurred due to the readers' noticing of the formality mismatch.

Following the same mechanism, I show the L2 participants' RT differences in Table 6.3.

Table 6.3

L2 participants' RT differences between the two conditions. F: formal, Inf: informal, C: context

Semantic	RT in		RT in		RT
Categories	formality-inconsistent		formality-consistent		difference
	condition (ms.)		condition (ms.)		(ms.)
Additive passages	InfAdd-FC:	182.563	FAdd-FC:	181.937	0.626
	FAdd-InfC:	160.810	InfAdd-InfC:	281.023	-120.213
Causal passages	InfCau-FC:	206.343	FCau-FC:	198.307	8.036
	FCau-InfC:	214.045	InfCau-InfC:	226.637	-12.592
Concessive passages	InfConce-FC:	139.780	FConce-FC:	181.063	-41.283
	FConce-InfC:	211.443	InfConce-InfC:	270.127	-58.684

On the basis of the hypothesis formed for this study, RT in the formality-inconsistent conditions was predicted to be longer than the RT in the formality-consistent conditions. Such a prediction, however, is clouded due to the generally longer reading time in the informal passages. This is particularly salient in those informal passages that include informal MDMs. The longer RT therefore leads to the negative RT differences observed in Table 6.3. Nevertheless, looking at the formal passages does reveal some facts about learners' pragmatic awareness. Other than the formal Concessive passage, we observe positive RT differences in both the Additive and the Causal passages, with 0.626ms in the former and 8.036ms in the latter case.

Two inferences can be drawn from the positive RT differences. It seems that L2 participants' noticing of the formality mismatch is more salient when reading formal passages with informal MDMs. Reading more formal compositions may be the reading mode that these L2 participants are familiar with. This finding correlates to the learner corpus finding that CLC learners notice the formal register and use MDMs accordingly. The second inference is that in terms of semantic types of MDM, L2 learners seem to notice the formality mismatch, particularly in Causal MDMs, shown by the positive RT difference of 8.039ms. Although L2 learners also notice the formality mismatch in passages with Additive MDMs, the pragmatic awareness raised is not as great as the case in the Causal MDMs, as shown by the small RT difference, 0.626ms.

The CLC corpus data shows that CLC learners produced the most Additive and Concessive MDMs. The formal-informal ratio analysis reveals that CLC learners' use of these two types of MDM in the formal register was less problematic. Using Additive and Concessive MDMs in the informal register remains a challenge. Causal MDMs, on the other hand, were found to be appropriate in both the formal and informal CLC writing. In other words, the two inferences drawn from Study 1 support the CLC corpus findings. L2 participants in Study 1 noticed the formality mismatch in the formal register and this is supported by the fact that CLC learners used MDMs more appropriately in the formal register. In terms of the three semantic types of MDM—Additive, Causal, and Concessive—L2 participants noticed the formality mismatch mostly in Causal MDMs, followed by Additive MDMs and then Concessive MDMs. This noticing order corresponds to the results observed in the CLC corpus.

6.2 Study 2 and the corpus evidence

6.2.1 A brief recap of Study 2

Study 2 was a multiple-choice discourse completion task (MDCT). The task involved ten passages, each with its linking MDM purposely omitted. Participants were asked to read the ten passages and to choose, from among ten provided MDMs, which would best fit the blanks. Like Study 1, the MDCT task was designed to test L2 learners' sensitivity to register difference by including two factors: semantics and formality. However, unlike Study 1, which tested only three semantic types of MDM, the MDCT task examined five semantic categories—Additive, Appositive, Causal, Concessive, and Enumerative. Both formal and informal passages were included. Combining the semantics and formality factors together, this means that five of the ten paragraphs were written in formal English (with five formal MDMs in the five semantic categories) and the other five passages in informal English (with five informal MDMs). In other words, each semantic category included a formal and an informal passage. For example, when reading the formal Causal passage, the most appropriate choice for the missing MDM is the formal Causal MDM *as a result*; semantically speaking, however, the informal Causal MDM *because of that* is also eligible.

The participants' data were examined via two measures, CLASS and EXACT MATCH, with the former focusing on the semantic factor, and the latter considering the formality factor. The CLASS measurement refers to whether or not participants have identified the original semantic type of MDMs. This means that for a formal passage that has its original Causal MDM, *as a result*, purposely omitted, both *as a result* and

because of that are legitimate choices, as both create the Causal link. The EXACT MATCH measure describes the situation when participants have identified the exact MDM that is used in the original passage. In a case like this, participants have not only identified the right class but also the correct register.

The hypothesis formed for Study 2 is that participants with a keen pragmatic awareness will notice not only the required semantic type of MDM but also the formality level of the passages as they read along. That is, participants' choice of a missing MDM will fulfil both the semantic linking requirement and the formality condition.

The L2 participants were eleven advanced (Adv) L2 learners and eighteen intermediate (Inter) L2 learners, with the former's average IELTS score being 7 to 7.5, and the latter 5.5 to 6.5. All of the twenty-nine L2 participants were native Chinese speakers. The IELTS score forms a bridge between the L2 participants and the CLC learners. The learner data archived in the CLC corpus can be sorted into six learner levels, based on these learners' IELTS scores. As IELTS provides a common ground for comparison, the Adv L2 participants' task performance will be compared mainly with the corpus data of C1 learners. The Inter L2 participants' task performance will be compared with the B2 learners. In addition, as the investigation of the learner corpus in Chapter 4 did not include Enumerative MDMs, the discussion in this section will not include Enumerative MDMs.

To present the discussion in a clear way, I will first provide a macro view of the experimental data and relate that to the corpus data. I will probe into the relationship

between the two measures and discuss any possible implications. Any inferences drawn from the comparison will be compared with the corpus data.

6.2.2 A macro view of the two measures and the corpus data

In Chapter 5, I have already shown the two groups' mean scores in the two measures as well as the differences between the two measures. Instead of repeating this information, I show the total mean scores of the five passages in the formal and informal register in Table 6.4.

Table 6.4

Total mean scores in the two registers by the two measures

CLASS measure	Adv L2	Inter L2
formal register	3.0910	2.2222
Informal register	2.2728	2.1666
EXACT MATCH measure		
formal register	1.7273	1.6111
Informal register	1.7273	1.6111

The total mean scores in the two registers of the CLASS and the EXACT MATCH measures provide valuable insights into how learners' noticing of register difference correlates to L2 learners' English proficiency. To begin with, the Adv L2 participants had a higher score than the Inter L2 participants in the two registers in both of the two measures, suggesting that English proficiency plays a role in identifying the right

semantic category as well as noticing the register difference. Even though the two groups had the same scores in the two registers under the EXACT MATCH measure, they both performed better in the formal register in the CLASS measure. This supports the CLC corpus finding, in which both C1 and the B2 learners showed a greater awareness in the formal register.

Another area of interest is the L2 participants' task performance in finding the correct semantic types of MDM. The two groups' mean scores in the CLASS measure are combined and presented on the basis of the register and class in Table 6.5.

Table 6.5

Summary of the L2 participants' mean scores (MS: mean score)

Semantic categories	MS in Formal register	MS in informal register
Additive	1.5404	1.5960
Causal	1.4495	1.0808
Appositive	1.3738	0.9899
Concessive	0.9495	0.7727

As the mean scores in Table 6.5 integrate the scores from both the Adv and the Inter L2 participants, they provide a global impression of how the L2 participants performed in the four semantic types of passage and in the two different registers. The ranking of the four passages from the highest to the lowest mean scores gives us an idea of how easy or difficult the passages are to the L2 participants. In the formal register, the ranking shows that while the L2 participants found identifying the

missing MDMs for the Additive and the Causal passages less challenging, they found the Concessive passage the most difficult. In the formal writing examined in the CLC corpus, I also discovered that the CLC learners used the Additive and the Causal MDMs in accordance with the formal context, but this is not the case with the Appositive or with the Concessive MDMs.

The ranking in the informal register shows a different order. Even though the L2 participants still found reading the Concessive passages difficult, their performance in identifying the Appositive cohesive tie was outstanding. The mean scores in the Additive and the Causal passages were in the middle. The informal writing investigated in the CLC corpus data shows that the CLC learners' use of Appositive MDMs was rather limited but they frequently made use of Additive MDMs, followed by the Concessive and the Causal MDMs. Although the ranking order in this study does not seem to correspond to the corpus finding, I discuss further the underlying reasons in Section 6.2.6.

6.2.3 Interpreting the relationship between the two measures

Although the two measures seem to focus on different aspects of the MDMs, one overlapping nature examined by both measures is semantics. This is because the EXACT MATCH measure identifies both the register and the semantics factors, whereas the CLASS measure only identifies the semantics factor. That is, the CLASS measure includes instances where participants fail in identifying the right register but succeed in identifying the required semantic type of MDM. Take for example the formal Causal passage mentioned earlier. Participants who select the formal MDM *as a*

result will be awarded one point under both the CLASS and the EXACT MATCH measure. Participants who select the informal MDM *because of that*, however, will only be awarded one point under the CLASS measure. That is, the scores in the EXACT MATCH measure are always smaller than the scores in the CLASS measure.

I have argued earlier (Section 5.4.6) that any differences observed between the two measures would quantify L2 participants' lack of awareness of the register difference. I will use L2 participants' mean scores in reading the formal Additive passages to illustrate this point. The Adv L2 participants' mean scores in selecting the appropriate MDM for this passage are 0.64 in the EXACT MATCH measure and 0.82 in the CLASS measure. This gives rise to a difference of -0.18 (0.64-0.82). The Inter L2 participants' mean scores in the two measures are 0.5 and 0.72 respectively and the difference is -0.22. While the majority of each group tended to notice the required Additive link (shown by the high mean scores in the CLASS measure), the Adv L2 participants showed a keener awareness of register difference. The Inter L2 participants' lack of formality awareness is evidenced by their greater measure difference, -0.22. In other words, the two measure differences here quantify the extent to which L2 participants are inattentive to the formality differences. One inference drawn from the relationship between the two measures is that a greater observed mean score difference suggests a more problematic pragmatic awareness.

In what follows, I present the mean score differences between the EXACT MATCH and the CLASS measure (referred to as two-measure differences) in the formal and the informal register and discuss the implications.

6.2.4 The two-measure differences in the formal register and the corpus data

Table 6.6 summarizes the two groups' two-measure differences in mean scores in the four conditions. The order of the four semantic conditions is ranked from the greatest two-measure difference to the least.

Table 6.6

L2 participants' two-measure differences in the formal conditions

Semantic categories	Adv L2	Semantic categories	Inter L2
Causal	-0.64	Causal	-0.33
Appositive	-0.46	Additive	-0.22
Additive	-0.18	Appositive	-0.05
Concessive	-0.09	Concessive	0
Sum	-1.37	Sum	-0.60

As suggested earlier, a greater two-measure difference indicates that participants are less attentive in the task. The scores in the sum rows show that the Adv L2 participants had problems in selecting formal MDMs for the passages in the formal register. Specifically, the Adv L2 participants had more problems in choosing the formal Causal and Appositive MDMs. This observation on the Causal passage contradicts the CLC corpus finding. The corpus data indicates that the C1 learners tended to use more formal Causal MDMs in both the formal and the informal register. That the Adv L2 participants in the MDCT task tended to choose the informal MDM *because of that* for the formal Causal passage may be explained by two reasons. First,

the selection of this passage might not be adequate. This is supported by the general tendency to select the informal Causal MDM for this formal passage found in both the NS and L2 participants (see Section 5.4.5.5). The second reason is that the participants' differing L1 backgrounds may influence their choice. All the L2 participants in Study 2 were Chinese learners of English, whereas the C- and the B-level learners in the CLC corpus were from various L1 backgrounds. Although this hypothesis needs to be tested, we cannot rule out the possibility that the L1 differences may be held accountable for the observed differences in the Causal passage.

The other passage that shows the Adv L2 learners' tendency to choose an informal MDM is the formal Appositive passage. Out of the nine Adv L2 participants who identified the correct Appositive class, five chose the informal MDM *what I'm saying is*. The CLC corpus data indicates that both the C1 and the B2 learners failed to distinguish between the formal and the informal register when using Appositive MDMs. In fact, the total number of Appositive MDMs retrieved in the CLC corpus was found to be the smallest, suggesting either an avoidance strategy or a lack of control in using Appositive MDMs appropriately. In other words, the L2 participants' task performance in the Appositive passage corresponds to the CLC corpus data.

In the Additive and the Concessive passages, both the Adv L2 and Inter L2 participants had smaller two-measure differences. Nevertheless, we can draw different inferences from this observation. In the case of the Additive passage, the two groups' greater mean scores in both the CLASS and the EXACT MATCH measure give rise to the small two-measure differences, indicating that both groups have not only identified the right class but also the right register. That said, there is a slight

difference between the two groups' task performance. Although the Inter L2 participants also showed a higher score in the CLASS measure, they had a lower score in identifying the right register. In other words, the Adv participants outperformed the Inter L2 participants in identifying the Additive link in the formal register. This performance difference is similar to the corpus finding. In the CLC corpus, in which I examined learners' writing in the formal register, I observed a gradual increase in the use of formal Additive MDMs as the learners' English proficiency develops.

In the case of the Concessive passage, although each of the two groups had small two-measure differences, the Adv L2 learners had higher scores in both the CLASS and the EXACT MATCH measure, whereas the Inter L2 participants had very low scores. The difference suggests that the Inter L2 participants had problems in identifying the Concessive semantic tie. The CLC corpus evidence also supports this finding. When compared with the C1 learners, the B2 learners in the CLC corpus used far fewer instances of formal Concessive MDMs in their formal writing, suggesting a lack of control in this regard.

In general, the Adv L2 participants performed better in the Additive and the Concessive passages in the formal register. Although they also had high scores in the CLASS measure in reading the Appositive and the Causal passages, the Adv L2 participants tended to select the informal MDMs for these two passages. This tendency is different from the CLC corpus finding and this will be further addressed in Section 6.2.6.

6.2.5 The two-measure difference in the informal register and the corpus data

The two-measure differences of the four passages in the informal register are shown in Table 6.7. Unlike the case in the formal register, where the Adv L2 participants had a higher sum score, the Inter L2 participants had a higher sum score in the informal register, suggesting that the Inter L2 participants tended to choose formal MDMs for the informal passages.

Table 6.7

L2 participants' two-measure differences in the formal conditions

Informal Register	Adv L2	Informal Register	Inter L2
Appositive	-0.18	Appositive	-0.17
Concessive	-0.18	Causal	-0.16
Additive	-0.09	Additive	-0.11
Causal	-0.09	Concessive	-0.11
Sum	-0.64	Sum	-0.55

Both groups of L2 participants had similar two-measure differences in choosing the missing MDM for the informal Appositive passage. In addition, the two groups also showed similar high mean scores in both the CLASS and the EXACT MATCH measures. This seems to suggest that both groups had noticed the informal register and had selected the informal MDM *what I'm saying is* accordingly. This observation, however, is not consistent with the CLC corpus data, where both the C1 and the B2 learners used very few instances of informal Appositive MDMs in the informal register.

In the Additive passage, although the two groups' two-measure differences are again rather similar, the Inter L2 participants had lower mean scores in both measures. This implies that many of the Inter L2 participants noticed neither the formality nor the semantic link required for this passage. The same phenomenon is found in the informal Causal passage, in which the two groups showed similar two-measure differences. The Adv L2 participants had higher mean scores in the CLASS and the EXACT MATCH measures yet again, indicating that the Adv L2 participants have a better grasp in using the informal Causal MDM *because of that* in the informal register.

The only passage in the informal register where the Adv L2 participants had lower mean scores than the Inter L2 participants in both measures was the Concessive passage. More Adv L2 participants chose the formal MDM *in spite of that/this* for the informal Concessive passage.

A tentative conclusion to draw from this discussion is that the Adv L2 participants show a heightened awareness of register difference and choose informal MDMs accordingly in reading the Additive, Causal, and Appositive passages. The only exception is the Concessive passage, for which the Inter L2 participants had higher mean scores in the two measures. This observation does not correspond to the CLC corpus data, in which the use of the informal Concessive MDMs was found to be quite limited in both the C1 and the B2 learners' informal writing. I elaborate more on this discrepancy in the following discussion section.

6.2.6 Discussion

In Chapter 4, where I examined the CLC corpus data, I concluded that, of the five semantic types of MDMs investigated, Causal and Contrastive are the least problematic, whereas Concessive and Appositive are the more challenging types for CLC learners. CLC learners seem to have a firm grip on using Additive MDMs in the formal register but not in the informal register. C-level learners were found to notice the register difference more than the B-level learners. In addition, the types and the frequencies of the MDM employed were correlated with the English proficiency of the L2 learners. The results of Study 2, however, do not seem to support the corpus evidence completely.

In the formal register of this MDCT task, the Adv L2 participants showed better task performance than the Inter L2 participants in the Additive and the Concessive passages. This concurs with the corpus finding, as the C1 learners used the Additive and the Concessive MDMs more appropriately in their formal writing. In the four informal passages, however, the Adv L2 participants performed better in the Additive and the Causal passages. The corpus data shows that the C1 learners outperformed the B2 learners in their use of the four semantic types of MDM in the informal register. In other words, I did not observe a strong correlation between changes in English proficiency on the one hand and changes in task performance on the other in Study 2. The inconsistency in this regard might be due to the very small number of L2 participants recruited in this study. In addition, unlike the C- and B-level learners, who had various L1 backgrounds, all the L2 participants in this task were Chinese speakers of English.

I have previously reported that the L2 participants' task performance in the Appositive passage in the informal register is different from the corpus data. I argue in what follows that this difference is explainable in light of the nature of the required task. In other words, despite the seemingly contradictory results, the corpus evidence and the experimental data complement each other to a certain extent.

The C1 and the B2 learners in the CLC corpus used very few instances of Appositive MDMs in either the formal or the informal register. Specifically, neither the C1 nor the B2 learners employed the informal Appositive MDM *what I'm saying is* in the corpus. Since not only the informal but also the formal Appositive MDMs were sparsely seen in the corpus, I have proposed that it is either that the CLC learners adopt an avoidance strategy, or that the concept of Appositive ties is beyond the reach of their understanding.

In the MDCT task, however, we observed a different pattern. When identifying the missing MDM for the formal Appositive passage, the Adv L2 participants showed a higher mean score in the CLASS measure even though their score in the EXACT MATCH measure was lower than the Inter L2 participants. That is, the Adv L2 participants selected the informal Appositive MDM *what I'm saying is* for the formal Appositive passage. In reading the informal Appositive passage, both the Adv and the Inter L2 participants had higher scores in the two measures, suggesting that identifying the Appositive MDM for an informal passage is certainly easier than the same task in the formal register.

Taking together the MDCT task result and the corpus evidence, I propose a plausible explanation in understanding L2 learners' use of Appositive MDMs. On the

one hand, the L2 participants' successful attempt to identify the Appositive cohesive tie in each of the registers indicates that these L2 participants are well aware of the Appositive semantic category. Their tendency to choose the informal Appositive MDM (*what I'm saying is*) in both the registers, on the other hand, shows that the use of the formal Appositive MDM (*that is to say*) might be a more challenging task and they therefore select the easier informal Appositive MDM by default. However, there exists an important attribute that differentiates the MDCT task from the CLC corpus data. While the MDCT is a receptive and semi-productive task, which is due to the fact that the participants are required to select an MDM, the writing that CLC learners compose is an absolute productive task. The decoding skill required in the MDCT task, comparatively speaking, is easier than the encoding skill required in the writing task. This explains that while the CLC learners tend not to use Additive MDMs in their writing, as this requires a higher level of mastery, the L2 participants in this study can easily identify the Appositive cohesive tie by selecting the informal Appositive MDM. On the face of it, the corpus evidence with regard to L2 learners' use of Appositive MDMs looks contradictory to the MDCT results; yet, by taking into account the task nature, the productive versus the receptive task, the L2 participants' performance in reading the Appositive passages does support the corpus finding.

Another interesting finding in the MDCT task is the L2 participants' scores in reading the Concessive passages. The Adv and the Inter L2 participants showed contrasting task performance. While the Adv L2 participants had high scores in each of the two measures in the formal register, the Inter L2 participants did better in the informal register. That the Inter L2 participants performed well in the informal

register is justifiable, as reading the informal passage is easier than reading the formal passage. The fact that the Adv L2 participants had a higher score in the formal register but a low score in the informal register, however, is less comprehensible. As is the case with the Causal passage, the informal Concessive passage might not be an ideal selection to be included in this task. This is supported by the twenty-nine NS participants' performance in the Concessive passages. While twenty-six of the twenty-nine NS participants identified the right class and the exact match of the MDM for the formal Concessive passage, only half of them successfully did so in the informal register.

6.3 The evaluation of the methodologies in the two studies

The methodologies employed in testing L2 learners' pragmatic awareness, the psycholinguistic experimental paradigm—reaction time task—as well as a discourse completion task, are not only new but also exploratory. Although these two methods are not unknown in SLA, my method of incorporating both in testing L2 learners' pragmatic awareness is new. On the basis of the comparisons and discussions made thus far, I will evaluate the plausibility as well as the applicability of the two methods.

The reaction time task involves participants' undivided attention and mental effort in task performance. Fatigue as well as the limitation of mental effort are possible factors that affect participants' performance. Because of this, I have only tested three semantic types of MDM. The RT task results show that while the NS participants showed a longer response time in reading formality mismatch sentences to some extent, the L2 participants did not. In other words, while the RT paradigm

reflects the NS participants' pragmatic awareness of register differences, it is not similarly indicative when the participants' awareness is less keen.

On the other hand, the MDCT task used in Study 2 seemed to detect both the NS as well as the L2 participants' awareness of register differences. Specifically, the Adv L2 and the Inter L2 participants showed different task performance, suggesting that English proficiency plays a role in L2 learners' ability to notice formality difference. Although Study 2 yields both supporting and contradictory experimental results when compared with the corpus data, the MDCT method seems to be more plausible and applicable than the RT task in this pursuit.

6.4 Summary and conclusion

In this chapter, I have compared the CLC corpus data with the two study results. Although the semantic types of the MDMs tested are slightly different, the results of the two studies have yielded some coherent findings, which support the CLC corpus evidence. In what follows I summarize the important findings and present them with the corpus evidence.

1. The L2 participants in both studies showed better performance in conditions in the formal register. This includes their greater noticing of the formality mismatch in the formal passages in the RT design, and their higher mean scores in reading the formal passages in the MDCT design. The learner corpus finding indicates that the CLC learners show better pragmatic awareness when using various MDMs in their formal writing.
2. Of the three semantic types of MDM tested in Study 1, L2 participants paid the

most attention to the formal Causal passages, followed by the formal Additive, and the formal Concessive passages. In the formal register of Study 2, the L2 participants also had a higher score in finding the missing MDMs for the Causal and Additive passages. They score the lowest in reading the Concessive passage. The CLC corpus data has shown that while the CLC learners' use of Additive and Causal MDMs in the formal register seemed to be less problematic, the CLC learners showed signs of struggling in using the Concessive and the Appositive MDMs.

3. The L2 participants showed a higher score in identifying the Appositive semantic tie; yet the CLC learners' use of Appositive MDMs was rather infrequent. This seemingly contradictory observation is due to the nature of the two different tasks. While the reading activity in the MDCT design was more of a receptive-oriented activity, the writing collected in the CLC corpus was a productive one. Looking at the L2 participants and the L2 writers' use of Appositive MDMs from this perspective, I regard the different findings as a reasonable expectation.
4. The MDCT task has proved to be an effective method in testing L2 pragmatic awareness. The predicted prolonged reaction time to be perceived in the RT task, on the other hand, might be too optimistic to be applied to L2 learners when their awareness of register difference is less sophisticated.

In the next chapter, I will elaborate on the insights gained from both the corpus studies and the two designed experiments. This is only an initial foray into L2 pragmatic awareness of register difference, yet with this groundwork laid, I will

conclude with the limitations of the methodologies adopted and set the agenda for future work.

Chapter 7 - Conclusion

In this chapter, I will first summarize the research that I have carried out in this thesis, followed by a discussion of both the applications and the limitations of this research. I conclude the thesis with possible directions for future work.

7.1 A summary of the thesis

This thesis started with my personal experience when I heard my classmate at the University of Cambridge, an advanced L2 learner, use the formal Causal marker *hence* in a casual conversation. That experience led to the question of appropriateness, which further evolved into my interest in investigating L2 learners' pragmatic awareness in using discourse markers (DMs) in various contexts. As single-word DMs, such as *hence* and *nevertheless*, have been examined extensively in the literature, I extended the notion of DMs to multi-word discourse markers (MDMs) and explored the possibilities of pinpointing L2 learners' pragmatic awareness in both the formal and the informal setting.

There were four main stages in this thesis. The first step was data mining in the native speaker (NS) corpora. The intuitively obvious assumption that a usage difference exists among various MDMs, such as the Concessive MDMs *in spite of that* and *but then again*, has to be verified by examining how these MDMs are used in the real world. I therefore incorporated four NS corpora—the BNC-Acad corpus, the BASE corpus, the Enronsent2 e-mail corpus, and the CANCODE corpus—in the verification process. The four corpora represent the formal written, formal spoken, informal

written, and informal spoken, respectively. I reviewed two methods that calculate the formality weight of DMs: the difference coefficient (*diff. coeff.*) and formality score (*FS*). Both methods were developed on the basis of the frequency information obtained from two corpora, a formal written corpus and an informal spoken one. Considering the dynamic nature of language, I included four corpora (rather than two) and devised the Weighted Formality Average method (*WFA*). I discovered that the intuitively clear impression of formal MDMs (*in spite of that*) and informal MDMs (*but then again*) could be confirmed by the results of *WFA* and the other two methods. Various MDMs can be placed along a formality continuum, depending on their individual formality weight. In other words, while some might argue that it is the semantic nuance that distinguishes *in spite of that* from *but then again*, it is in fact the formality weight that distinguishes one from the other in usage difference. Such a usage difference in turn reflects the formality of the context within which these MDMs occur.

After establishing the formality differences for five semantic types of MDM, the next stage was data mining in the learner corpus. I examined the same set of MDMs that was searched in the NS corpora in the Cambridge Learner Corpus (CLC). As the user interface of the CLC corpus offers L2 writer metadata, I was able to investigate not only how L2 learners' English proficiency affected the quantity and types of the MDMs used, but also how L2 learners employed these MDMs in both formal and informal writing. The results suggested that the advanced C-level learners are more linguistically resourceful, employing a variety of MDMs, particularly evidenced by more semantic types and formality types of MDM retrieved from the C2 learner data.

Among all the five semantic types of MDM explored, CLC learners were found to employ Additive and Concessive MDMs frequently. Appositive MDMs were found to be the least employed. Taking into account the formality of the context, I discovered that CLC learners were capable of using the Additive MDMs in the formal register but they overused the formal, Additive MDMs in the informal register. While they did not show particular register problems in using Causal and Contrastive MDMs in formal and informal settings, CLC learners were relatively incompetent in using Concessive and Appositive MDMs appropriately in each of the two registers.

The third stage was to test L2 learners' pragmatic awareness in using MDMs via experiments, the purpose of which was to investigate whether or not L2 learners' pragmatic awareness is perceivable. The two experiments that I designed were a reaction time task and a multiple-choice discourse completion task (MDCT). Both studies were designed to include a formality mismatch or inconsistency: i.e., the formality of the MDMs and the context did not match. The design rationale was that L2 participants with developed pragmatic awareness would notice the embedded formality mismatch, and their choice of MDMs would correspond to the formality of the context. In Study 1, the NS participants showed their pragmatic awareness with a prolonged reaction time when reading the formality-inconsistent sentences. The L2 participants, on the other hand, did not show any hesitation. In the MDCT task, both the NS and advanced L2 participants showed a keener awareness of the formality difference, whereas the intermediate L2 participants did not notice the register difference.

The last stage of this thesis was to compare the learner corpus findings with the results of the two studies. The purpose was to examine whether or not the experimental findings would support the results obtained from the learner corpus. For example, would the MDMs that L2 participants paid the most attention to correspond to the types of MDM that were found to be associated with the fewest problems in the CLC corpus? The comparisons showed that the experimental results were consistent with the learner corpus data, although some incompatible areas were also observed. The consistent parts referred to the CLC learners' and the L2 participants' performance in Additive and Concessive MDMs. While CLC learners showed a firm grasp of the use of Additive and Causal MDMs, the L2 participants' RT in Study 1 also showed their awareness of the formality-inconsistent condition in the Causal passages by their longer response time. Furthermore, the L2 participants in Study 2 had the highest mean scores in identifying the Additive cohesive tie. CLC learners showed a lack of register awareness when using the Concessive MDMs in their writing. Likewise, the majority of the L2 participants were also unsuccessful in noticing the formality-inconsistency in the Concessive passages. Incompatible results were found in the Causal passages. While CLC learners used Causal MDMs appropriately by paying attention to the formality of the context, the L2 participants in Study 2 tended to choose the informal MDM (*because of that*) for the formal Causal passage. As explained in Chapter 6, the inconsistency might be due to the L1 background difference or the inappropriate selection of that passage. The CLC learners were a large group of L2 learners with various L1 backgrounds whereas the L2 participants in the MDCTs task were mainly Chinese learners of English.

7.2 General conclusions

The learner corpus data has shown that L2 learners face a challenge in using MDMs in accordance with register differences. Nevertheless, instead of drawing a conclusion based on an overuse/underuse dichotomy, this thesis has shown that the lack of register awareness that leads to overuse or underuse phenomenon has to be discussed by taking into account the semantic types that MDMs represent. The two experiments, which tested various semantic types of MDMs by implanting a formality mismatch in the reading passages, have also shown that the L2 participants' pragmatic awareness varies depending on the semantic types and registers of passages they read. In other words, the approach taken in evaluating the learner corpus data and the two experimental results has shown that there exists a hierarchy of acquisition difficulty in the five semantic types of MDM. While the L2 learners did not show a problem in employing the Causal and Additive MDMs, they have difficulties in using Concessive and Appositive MDMs appropriately.

One of the issues in examining corpus data is missing data, which, when not acknowledged, can "introduce bias into the models being evaluated and lead to inaccurate data mining conclusions" (Brown & Kros, 2003). For example, the Appositive MDMs were used sparingly in both the formal and informal writing in the CLC corpus. As discussed in Chapter 4, such an underuse phenomena might be indicative of an avoidance strategy that CLC learners adopt, suggesting that L2 learners find the Appositive concept a much more complex notion. Following this line of reasoning, a legitimate prediction of the experimental result would be that the L2

participants might have the lowest mean score in identifying the Appositive link in the MDCT experiment. Contrary to this expectation, the MDCT result showed that the L2 participants had a very high mean score in identifying the missing Appositive link. As the MDCT design is a receptive task and the CLC learner writing a productive one, we might argue that the different results come from the difference in the nature of the two tasks. Because of the different natures of the two tasks, one type of data indeed complements the other. The two types of data have shown that while the Appositive concept might be easy to comprehend (as shown by the high mean scores in the MDCT experiment), the use of Appositive MDMs in writing is beyond L2 learners' reach. In other words, the MDCT experiment has provided an alternative channel for investigating L2 learners' understanding of the Appositive concept. Through the comparison of the two types of data (corpus data vs. experimental data), this thesis has provided a feasible solution to the missing data problem (Appositive MDMs).

The devised *WFA* method, which calculated the formality weight of MDMs by considering frequencies from four corpora, might suffer from the criticism of involving a more complicated calculation when compared with Altenberg's *diff. coeff.* and Brooke et al.'s *FS* method. Nevertheless, the comparison of the three methods has demonstrated that the *WFA* method offers a more objective approach in showing the formality weight of MDMs in relation to the frequency factor. An example is the formality weight of *add to this*. The *diff. coeff.* and *FS* methods gave scores of -0.81 and -0.60 to this MDM, suggesting that it carries a very formal weight. The *WFA* method, however, only assigned this MDM a score of -0.22 because *add to this* occurred with relatively low frequencies across the four corpora. That is, the *WFA* method showed

the impact that frequency had upon formality weight. The formality continuum created by the *WFA* method, therefore, reflects a less rigid representation of formality weight.

7.3 Limitations

This thesis has adopted a new methodology, which includes both corpus-based studies and the devised experiments, in its attempt to show L2 pragmatic awareness. The new approach has limitations in two respects. The first problem results from the corpus-based studies, and the second is related to the two experiments.

The first limitation lies in the four corpora used in creating the formality continuum (Chapter 3). While three NS corpora archive British English data, the Enronsent2 e-mail corpus comprises messages written mainly in American English. The rationale for including an e-mail corpus is based on the assumption that the creation of the formality continuum should be done by taking language samples from various contexts, showing the idiosyncratic nature of the language itself in different registers. The lack of a large e-mail corpus in British English led to the inclusion of the Enronsent2 corpus. The extent to which the difference between British and American English may affect language users' choices of MDMs is uncertain. Therefore, the inclusion of the American English corpus with the other three British English corpora is a compromise in this pursuit. Nevertheless, the American vs. British English issue is yet to be resolved.

Another corpus-related limitation is the comparison made among various L2 learners in the CLC corpus. In the formal writing of the CLC corpus, four different

sub-corpora were created based on the L2 learners' English levels—C2, C1, B2, and B1. In the informal register, however, only three sub-corpora were available—C1, B2, and B1. In other words, a direct comparison between the C2 learners' use of the MDMs in both the formal and the informal register is missing. Had the informal C2 sub-corpus been available, a more complete description of the CLC learners' use of MDMs would have been offered.

The second limitation with regard to the thesis is due to the methodology in the two studies. Study 1 made use of the reaction time measurement, with the purpose of seeing whether awareness of register difference with regard to the use of MDMs is perceivable via a prolonged reaction time. To my knowledge, this was the first study that incorporated reaction time toward such a purpose. Although the NS participants showed longer reaction times in reading style-inconsistent sentences, the L2 participants did not show such a tendency. While the result of this study has proven its feasibility, reflecting L2 pragmatic awareness might require a more careful design.

Study 2 used the discourse completion task design, which is conventionally used to test various speech acts in L2 learners. Unlike the results of Study 1, both the NS and the advanced L2 participants' pragmatic awareness was pinpointed in the multiple-choice discourse completion task (MDCT). Nevertheless, each of the studies suffers from the same problem: the use of corpus-informed materials in the two designs. Adopting corpus-informed materials provides an authentic language input, which in turn is expected to elicit the most natural answers from the participants. As one basic idea of using corpus material is its authenticity, such material is often used as it is, without further editing. Providing authentic materials (colloquial English in the

written form and very formally written passages) for the L2 participants in the two studies, however, led to some comprehension problems. Extra care or steps are required when presenting corpus-informed materials to L2 learners. Employing corpus-informed materials with a pre-editing step might yield more fruitful experimental results; however, the amended texts might suffer from a partial loss of originality.

7.4 Directions for further research and its application

Despite the limitations mentioned, the results of the corpus-based studies as well as the two experiments have yielded important insights into how L2 learners show their pragmatic awareness when using various MDMs in two registers. Nevertheless, the aforementioned limitations suggest that there is room for improvement. In what follows, I will discuss three directions for future research: methodology, learner corpus data, and types of learner.

7.4.1 Methodology

Three aspects of methodology will be addressed: the corpus-based approach in creating the formality continuum, the framework used to examine learners' pragmatic awareness, and the experimental approach in testing L2 pragmatics.

This thesis has shown that the inclusion of four English NS corpora in examining the formality weight of MDMs is a plausible approach. This method can be applied in examining formality weight of other types of multi-word unit, such as phrasal verbs or idioms. Take *run a risk* for example. Other possible synonymous idioms include *take a*

chance, *take a risk*, *risk one's neck*, and *take an opportunity*. L2 learners are likely to treat these synonymous units as interchangeable phrases. By applying the WFA method to create the formality weight of these synonymous units of *run a risk*, I suggest that we will be able to offer L2 learners context-sensitive language use, showing how register differences affect linguistic choices.

The framework used in examining L2 English in the CLC corpus has shown that word frequency counts can inform more than the digits they represent. Frequency counts in learner corpus studies tend to be compared with the frequencies obtained from NS corpora. One common conclusion drawn from such comparisons is the phenomenon of L2 learners' overuse or underuse. Taking into account the register differences as well as the semantic types of the MDMs, this thesis has demonstrated that the quantitative analysis of the frequency count, when combined with the information on the register difference, L2 learners' proficiency level, and the semantic types of MDMs, can in turn be used as the information needed for a qualitative analysis. Such a framework can be applied to other types of learner corpus data, especially when the lack of register awareness is a major concern. An example is L2 learners' use of colloquial words in collocations, such as *big influence*, which is found to occur in the CLC corpus eighty-nine times, or 1.9 times per million words. By applying this framework to analyse learners' awkward collocations like *big influence* and its more appropriate versions, such as *great influence*, *profound influence*, *huge influence*, and *strong influence*, we will be able to identify in what contexts and at what levels of English proficiency L2 learners tend to use this collocation, rather than treating the word *big* in *big influence* as a word choice problem.

The third direction of further research in methodology is the experimental approach. The NS participants showed their pragmatic awareness in formality mismatch conditions in the experiments. Their awareness is particularly noticeable in the MDCT task. Only the advanced L2 participants showed their awareness in this regard. As the noticing of register difference forms part of readers' general impression of the reading passages, it seems that the intermediate L2 participants' attention flow simply focused on the local sentence levels in the task. In other words, including L2 learners with higher English proficiency, or bilinguals, such as Chinese-English bilinguals, as participants might yield interesting results.

7.4.2 Learner corpus data

This thesis has used the CLC corpus data, which is a database of L2 writing collected from EFL (English as a Foreign Language) learners with various L1 backgrounds. Although the metadata of the L2 writers (gender, L1, English levels) and the writing (formal, informal writing) makes possible the comparison of two registers and learners' use of MDMs at various English levels, the CLC corpus only includes one type of L2 writing: examination scripts. The sole data source makes it difficult to equate the data analysis with the entire body of L2 learners. One of the future research directions is to include other learner corpora.

7.4.3 Types of learner

The MDCT task has proved to be a plausible method for testing L2 pragmatic awareness. Another direction for future work is to replicate the study by expanding

the number of L2 participants with different L1 backgrounds and various English levels. Specifically, two types of learner should be included: L2 learners who acquire English differently and L2 learners who have different L1 backgrounds.

Although the L2 writers of the CLC corpus come from various L1 backgrounds, these L2 learners are homogeneous in the way they acquire English, which is through a classroom setting. Another group of English learners, the English as a Second Language (ESL) learners, acquire English not only in classrooms but also from their environment where English is the main or official language. The ESL learners presumably receive different kinds of input from the EFL learners. As research has suggested that the amount of input affects the development of L2 learners' pragmatic competence (Kasper & Roever, 2005), one question worthy of investigating is whether or not ESL L2 learners will develop a keener pragmatic awareness than will EFL L2 learners.

The other direction of future research in *types of learner* is L2 learners' L1 factor. The L2 writers in the CLC corpus were analysed according to their English proficiency levels. The influence that their L1 background might have on their choices of MDMs in various registers, on the other hand, was not further investigated. One interesting finding that might be related to the L1 background was observed in the MDCT task. Both the advanced and the intermediate L2 participants showed an identical ranking order in the mean scores they obtained in the five semantic types of passages, with the highest scores for the Appositive passages and the lowest for the Concessive passages (see Section 5.4.6). Nevertheless, the extent to which the L1 factor of these L2 participants has influenced their task performance was not further pursued. Focusing

on L2 learners with a particular L1 background offers a straightforward comparison between the corpus evidence and the experimental results. For example, analysing the CLC data produced by L2 writers with a Chinese L1 background, we will be able to make a direct comparison between the corpus data and the Chinese L2 participants in the MDCT task, providing more insights into how a particular group of L2 writers (in this case, Chinese learners of English) develop their pragmatic awareness in using MDMs.

7.5 Final remarks

Pragmatic awareness, unlike verbal behaviour or written production, is not easy to perceive; yet, it is one of the factors that determines how a language speaker responds in various linguistic contexts. Using the analogy of an iceberg to explain the interrelationship between the less-discernible pragmatic awareness and the observable linguistic choices, the data obtained from the CLC corpus function like the tip of the iceberg, whereas the devised Studies 1 and 2 serve as a *magnifier* to reveal what is hidden beneath: the pragmatic awareness. Certainly there are more invisible parts of the iceberg that await discovery. This thesis has made a first attempt at quantifying L2 learners' pragmatic awareness in registers on the basis of the corpus and the experimental data. With the rise of research interest in Interlanguage Pragmatics, this thesis adds its contribution to the current research and brings a deeper understanding of L2 pragmatic awareness.

References

- Achiba, M. (2003). *Learning to Request in a Second Language. A Study of Child Interlanguage Pragmatics*. Clevedon: Multilingual Matters.
- Adolphs, S. (2008). *Corpus and Context: Investigating Pragmatic Functions in Spoken Discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55-76). Amsterdam: John Benjamins.
- Alcón, E., & Jórda, M.P.S. (2008). Pragmatic awareness in second language acquisition. In Cenoz, J. & Hornberger, N. (Eds.), *Encyclopedia of Language and Education*, vol 6, (pp. 193-204). New York: Springer.
- Altenberg, B. (1986). Contrastive linking in spoken and written English. In G. Tottie and I. Backlund (Eds.), *English in Speech and Writing* (pp. 13-40). Uppsala: Semqvist & Wiksell International.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101-122). Oxford : Oxford University Press.
- Altenberg, B. & Tapper, M. (1998). The Use of Adverbial Connectors in Advanced Swedish Learners' Written English. In S. Granger (Ed.), *Learner English on Computer* (pp. 80-93). London, New York.
- Andersen, E. S. (1990). *Speaking with Style. The Sociolinguistic Skills of Children*. London: Routledge.

- Arapoff, N. (1968). The Semantic Role of Sentence Connectors in Extra-Sentence Logical Relationships, *TESOL Quarterly*, Vol. 2, No. 4 (Dec., 1968), pp. 243-252
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Balkin, J.M. (1994). Being Just with Deconstruction, *Social Legal Studies*, 3 (3), 393-404.
- Bardovi-Harlig, K. (1999). Exploring the interlanguage of interlanguage pragmatics: A research agenda for acquisitional pragmatics. *Language Learning*, 49, 677-713.
- Bardovi-Harlig, K., & Dörnyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. In: *TESOL Quarterly*, 32 (2), 233-262.
- Bardovi-Harlig, K., & Hartford, B. (1993). Learning the rules of academic talk: A longitudinal study of pragmatic development. *Studies in Second Language Acquisition*, 15, 279-304.
- Bardovi-Harlig, K., & Hartford, B. (2005). *Interlanguage Pragmatics. Exploring Institutional Talk*. Mahwah, NJ: Erlbaum.
- Bardovi-Harlig, K., & Mahan-Taylor, R. (2003). *Teaching Pragmatics*. Washington, DC: United States Department of State.
- Baron, N. S. (2003). Why email looks like speech: Proofreading, pedagogy and public face. In J. Aitchison & D. M. Lewis (Eds.), *New media language* (pp. 85-94). New York: Routledge.
- Barron, A. (2002). *Acquisition in IL Pragmatics: How to Do Things With Words in a Study Abroad Context*. Amsterdam: John Benjamins.

- Barton, E. L. (1993). Evidentials, argumentation, and epistemological stance, *College English*, 55 (7), 745-769.
- Bell, D. M. 2004. "Correlative and non-correlative on the other hand". *Journal of Pragmatics*, 36 (12), 2179-2184.
- Bell, D. M. (2010). Nevertheless, still and yet: Concessive cancellative discourse markers, *Journal of Pragmatics*, 42 (7), 1912-1927.
- Berkenkotter, C.& Huckin, T. (1993). Rethinking genre from a sociocognitive perspective, *Written Communication*, 10, (4), 475-509
- Bialystok, E. (1993). Metalinguistic awareness: The development of children's representations of language. In C. Pratt & A. Garton (Eds.), *Systems of Representation in Children: Development and Use* (pp. 211-233). London: Wiley & Sons.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1994). An analytical framework for register studies. In D. Biber & E. Finegan (Eds.), *Sociolinguistic Perspectives on Register* (pp. 31-56).
- Biber, D. (2006). *University Language. A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins Publishing Company.
- Biber, D. & Conrad S. (2001). Register Variation: A Corpus Approach, In D. Schiffrin, D. Tannen & H. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 175-96). Oxford: Blackwell.
- Biber, D., Conrad, S. & V. Cortes. (2004). If You Look at...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics* 25, 371-405.

- Biber, D. & Finegan, E. (1994). *Sociolinguistic Perspectives on Register*. Oxford: Oxford UP.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Bikeliene, L. (2008). Resultive Connectors in Advanced Lithuanian Learners' English writing. *Kalbotyra*, 59 (3), 30-37.
- Birner, B. (1988). *But* as conventional implicature: Identifying the source of contrast. In *Northwestern University Working Papers in Linguistics*, 1, 16-28.
- Blakemore, D. (1987). *Semantic Constraints on Relevance*, Blackwell, Oxford.
- Blakemore, D. (1988). 'So' as a constraint on relevance. In R. Kempson (Ed.), *Mental Representations. The interface between language and reality*. Cambridge: Cambridge University Press.
- Blakemore, D. (1992). *Understanding Utterances*, Blackwell.
- Blakemore, D. (2000). Indicators and procedures: *nevertheless* and *but*. *Journal of Linguistics*, 36, 463-486.
- Blakemore, D. (2002). *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*, Cambridge University Press.
- Blass, R. (1990). *Relevance Relations in Discourse: A Study with Special Reference to Sissala*. Cambridge: Cambridge University Press.
- Bolden, G. B. (2009). Implementing incipient actions: The discourse marker 'so' in English conversation, *Journal of Pragmatics*, 41, 974-998.
- Bolinger, D. (1972). *That's That*. The Hague: Mouton.

- Bouton, L. F. (1992). The interpretation of implicature in English by NNS: Does it come automatically—without being explicitly taught? In L.F. Bouton & Y. Kachru (Eds.), *Pragmatics and language learning* (Vol. 3, pp. 53-65). Urbana-Champaign: University of Illinois, Division of English as an International Language.
- Britton, B. K., Glynn, S. M., Meyer, B. J. F., & Penland, M. J. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology*, 74, 51-61.
- Brizuela, M., Andersen, E. & Stallings, L. (1999). Discourse Markers as Indicators of Register, *Hispania*, Vol. 82, No. 1, pp. 128-41.
- Brooke, J., Wang, T., & Hirst, G. 2010. "Automatic acquisition of lexical formality". *COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics*, 90-98.
- Brizuela, M., Andersen, E. & Stallings, L. 1999. „Discourse Markers as Indicators of Register". *Hispania*, Vol. 82, No. 1, 128-41.
- Brown, M.L., & Kros, J. F. (2003). The impact of missing data on data mining. In J. Wang (Ed.), *Data mining: opportunities and challenges* (pp. 174-198). IGI Publishing Hershey, PA.
- Bunz, U., Campbell, S.W. (2004). Politeness Accommodation in Electronic Mail, *Communication Research Reports*, 21 (1), 11-25.
- Buzzanell, P., Burrell, N., Stafford, S., & Berkowitz, S. (1996). When I call you up and you're not there: Application of communication accommodation theory to telephone answering machinemessages. *Western Journal of Communication*, 60, 310.
- Campbell, R., & Wales, R. (1970). The study of language acquisition. In J. Lyons (Ed.), *New horizons in linguistics* (pp. 242-260). Harmondsworth: Pelican.

- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication*. New York: Longman.
- Canale, M. and Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1 (1), 1-47.
- Carter, R. A. and M. J. McCarthy. (1995). Grammar and the spoken language. *Applied Linguistics*, 16 (2), 141-158.
- Carter, R. & McCarthy, M. (2004). This that and the other: Multi-word clusters in spoken English as visible patterns of interaction, *TEAGNA*, 21, 30-52.
- Carter, R. A. and M. J. McCarthy. (2006). *Cambridge Grammar of English: A Comprehensive Guide to Spoken and Written Grammar and Usage*. Cambridge: Cambridge University Press.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6 (2), 5-35.
- Chafe, W.L. (1985). Linguistic differences produced by differences between speaking and writing, In D.R. Olson, A. Hildyard, & N. Torrance (Eds.), *Literacy, language, and learning: The nature and consequences of reading and writing*, (pp.105-23).
- Chalker, S. & Weiner, E. (1994). *The Oxford dictionary of English grammar*, Oxford, Oxford University Press.
- Chaudron, C. & Richards, J. C. (1985). The Effect of Discourse Markers on the Comprehension of Lectures, *Applied Linguistics*, 7(2), pp.113-127.

- Chen, C. W. (2006). The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learners. *International Journal of Corpus Linguistics* 11 (1), 113-130.
- Cheng, X., & Steffensen, M.S. (1996). Metadiscourse: A technique for improving student writing. *Research in the Teaching of English*, 30, 149-181.
- Christie, C. (2000). *Gender and Language: Towards a Feminist Pragmatics*. Edinburgh: Edinburgh University Press.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Cohen, A., & Olshtain, E. (1993). The production of speech acts by EFL learners. *TESOL Quarterly*, 27, 33-56.
- Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In J. McH. Sinclair (Ed.), *How to use Corpora in Language Teaching* [Studies in Corpus Linguistics 12] (pp. 67-85). Amsterdam: John Benjamins.
- Cortes, V. 2004. Lexical Bundles in Published and Student Writing in History and Biology. *English for Specific Purposes* 23 (4), 397-423.
- Coupland, N. (2007). *Style: Language Variation and Identity*, Cambridge university press.
- Coupland, N. (2010). Accommodation theory, In J. Jaspers, J. Östman & J. Verschueren (Eds.), *Society and Language Use* (pp. 21-27). Amsterdam: John Benjamins
- Crewe, W. J. (1990). The illogic of logical connectors. *ELT Journal*, 44, 316-325.
- Crismore, Avon, Markkanen, R. & Steffensen, M. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication*, 10 (1), 39-71.

- Crystal, D. (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Davis, T.C., Wolf, M.S., & Bass, P.F.(2006). Literacy and misunderstanding of prescription drug labels. *Ann Intern Med*.145, 887– 894.
- De Felice, F., & Deane, P. (2009). *Identifying speech acts in emails: Business English and non-native speakers*. Corpus Linguistics, Liverpool, July 2009.
- Dewaele, J. (2008). "Appropriateness" in foreign language acquisition and use: Some theoretical, methodological and ethical considerations, *International Review of Applied Linguistics*, 46 (3), 235-255.
- Dredze, M., Hanna, W., Danny, P., Tova, B., Josh, C., & Joshua, M. 2008. "Intelligent email: Aiding users with AI". NECTAR Paper. *Proceedings of American National Conference on Artificial Intelligence (AAAI)*, 23, 1524-1527.
- DuFon, M. A. (1999). *The acquisition of linguistic politeness in Indonesian as a second language by sojourners in a naturalistic context*. (Doctoral dissertation, University of Hawaii.) Dissertation Abstracts International, 60, 3985.
- DuFon, M. A. (2000). 'The acquisition of negative responses to experience questions in Indonesian as a second language by sojourners in naturalistic interactions,' in B. Swierzbin, F. Morris, M. Anderson, C.A. Klee, and E. Tarone (eds.), *Social and Cognitive Factors in Second Language Acquisition*. Somerville, MA: Casadilla Press. Pp. 77-97.
- Duranti, A. (1985). Sociocultural dimensions of discourse. In T. van Dijk (Ed.), *Handbook of Discourse Analysis*, (Vol. 1, pp. 193-230). London: Academic Press.

- Ellis, R. (1992). Learning to communicate in the classroom: A study of two language learners' requests. *Studies in Second Language Acquisition*, 14 (1), 1-23.
- Ellis, J. & Ure, J. (1969). Language Varieties: Register. In R. Meachan (Ed.), *Encyclopedia of Linguistics: Information and Control* (pp. 251-259). London: Pergamon.
- Félix-Brasdefer, J.J. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community, *Language Learning*, 54 (4), 587-653.
- Ferguson, C. (1977). Babytalk as a Simplified Register. In C. Snow & C. A. Ferguson (Eds.), *Talking to Children* (pp. 209-235). Cambridge: CUP.
- Field, Y. & Yip, L.M.O. (1992). A comparison of internal cohesive conjunction in the English essay writing of Cantonese speakers and native speakers of English. *REL C Journal*, 23 (1), pp.15-28.
- Finegan, E. (2008). *Language: Its Structure and Use*, 5th ed. Thomson/Wadsworth.
- Fitzpatrick, T. (2005) An exploration of factors affecting the acquisition and the use of formulaic sequences in the L2. In Cosme, C., Gouverneur, C., Meunier, F. & Paquot, M. (eds) *Proceedings of the Phraseology 2005 Conference*, Louvain-la-Neuve, 13-15 October 2005, 137-139.
- Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, 38, 19-33.
- Fraser, B. (1990). An Approach to Discourse Markers. *Journal of Pragmatics*, 14, 383-95
- Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6 (2), 197-190.

- Fraser, B. (1998). Contrastive Discourse Markers in English, In A. H. Jucker & Y. Ziv (Eds.), *Discourse Markers: Descriptions and Theory* (pp. 301-326). Amsterdam: John Benjamins.
- Fraser, B. (2009). The English contrastive discourse marker on the contrary, In Fraser, B. & Turner, K. (Eds.), *Language in Life, and a Life in Language: Jacob Mey-A Festschrift* (pp. 87-96), Bingley: Emerald Group Publishing Limited
- Fung, L. & Carter, R. (2007). Discourse Markers and Spoken English: Native and Learner Use in Pedagogic Settings, *Applied Linguistics*, 28 (3), 410-439.
- Gale, W.A. & Church, K.W. 1994. "What's wrong with adding one?" In N. Oostdijk & P. de Haan (Eds.), *Corpus-Based Research into Language*. Rodolpi, Amsterdam.
- Gale, W.A. 1995. "Good-Turing smoothing without tears". *Journal of Quantitative Linguistics*, 2, 217-237.
- Garcia, P. (2004). Developmental Differences in Speech Act Recognition: A Pragmatic Awareness Study, *Language Awareness*, 13 (2), 96- 115.
- Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, 15, 87-105.
- Giles, H. & Coupland, N. (1991). *Language: Contexts and Consequences*. Keynes: Open University Press.
- Giles, H., Coupland, J. & Coupland, N. (1991). *Contexts of accommodation: developments in applied sociolinguistics*. Cambridge University Press.
- Giles, H., Taylor, D.M., and Bourhis, R. (1973). Towards a Theory of Interpersonal Accommodation through Language: Some Canadian Data, *Language in Society*, 2 (2), 177-192.

- Giles, H., Willems, M., Gallois, C., and Anderson, M.C. (2007). Accommodating a new frontier: The context of law enforcement. In Fiedler, K. (Ed.). *Social communication*. New York: Psychology Press.
- Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation, 2008, *English Text Construction*, 1 (1), 41-61
- Gnisci A. (2005). Sequential strategies of accommodation: a new method in courtroom, *British Journal of Social Psychology*, 44, 621-643.
- Granger S. (1993). The International Corpus of Learner English. *The European English Messenger*, 2 (1), 34.
- Granger S. (1994). The Learner Corpus: A Revolution in Applied Linguistics. *English Today*, 39 (10/3), 25-29.
- Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English, *World Englishes*, 15 (1), 17-21.
- Granger S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In Cowie A. (Ed.), *Phraseology: theory, analysis and applications* (pp. 145-160). Oxford: Oxford University Press.
- Granger, S. & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English, *World Englishes*, 15 (1), 17-21.
- Greenbaum, S. (1996). *The Oxford English Grammar*, Oxford: OUP.
- Gries, S. 2008. "Dispersion and Adjusted Frequencies in Corpora". *International Journal of Corpus Linguistics*, 13 (4), 403-437.

- Gries, S. 2011: online. Useful statistics for corpus linguistics. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.160.9846&rank=1> (accessed Aug 2011).
- Grosz, B.J. and Sidner, C.L. (1986). Attention, intentions and the structure of discourse, *Computational Linguistics*, 12, 175–204.
- Hess, C.W., Landry, R.C., & Sefton, K.M. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech and Hearing Research*, 29, 129-134.
- Halliday, M.A.K. (1985). *Spoken and Written Language*, OUP.
- Halliday, M.A.K. & Hasan, R. (1976). *Cohesion in English*. London: Longman
- Hassall, T.J.: (1997). *Requests by Australian Learners of Indonesian*, Unpublished doctoral dissertation, Australian National University, Canberra.
- Halliday, M.A.K., McIntosh, M., & Stevens, P. (1964). *The linguistic sciences and language teaching*. London: Longman.
- Hatch, E. (1992). *Discourse and Language Education*. Cambridge: Cambridge University Press.
- Herring, S.C. & Paolillo, J.C. 2006. "Gender and genre variation in weblogs". *Journal of Sociolinguistics*, Vol 10 (4), 439-459. doi: 10.1111/j.1467-9841.2006.00287.x
- Heylighen, F. & Dewaele J.-M. (1999). Formality of Language: definition, measurement and behavioural determinants. Internal Report, Center "Leo Apostel", Free University of Brussels.
- Heylighen, F. and Dewaele, J.-M. (2002). Variation in the contextuality of language: An empirical measure. *Foundations of Science*, 7, 293-340.

- Hinkel, E. (1999). Objectivity and credibility in L1 and L2 academic writing. In: Hinkel, E. (Ed.), *Culture in Second Language Teaching and Learning*. Cambridge University Press, Cambridge, pp. 90–108.
- Hinkel, E. (2001). Matters of Cohesion in L2 Academic Texts, *Applied Language Learning*, 12, 111-132.
- Hinkel, E. (2002). *Second language writers' text*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hinkel, E. (2004). *Teaching Academic ESL Writing. Practical Techniques in Vocabulary and Grammar*. Lawrence Erlbaum Associates.
- Hofland, K. & Johansson, S. (1982). *Word frequencies in British and American English*. Bergen: The Norwegian Computing Centre for the Humanities.
- Holmes, J. (1993). New Zealand women are good to talk to: an analysis of politeness strategies in interaction. *Journal of Pragmatics*, 20, pp.91-116.
- Hovy, E. & Maier, E. (1995). Parsimonious or profligate: How many and which discourse relations? Technical report, University of Southern California.
- Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, 30, 437–455.
- Hyland, K. (2004). Disciplinary interactions: metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13, 133–151.
- Hyland, K. (2008). As Can Be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*, 27, 4-10.
- Hylan, K. & Tse, P. (2004). Metadiscourse in academic writing: a reappraisal, *Applied Linguistics*, 25 (2), 156-177.

- Hymes, D. (1972). Communicative competence. In J.B. Price & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin, 269-293.
- Intaraprawat, P. & Steffensen, M. (1995). The use of metadiscourse in good and poor ESL essays, *Journal of Second Language Writing*, 4, (3), pp.253-272.
- Irvine, J.T. (1979). Formality and Informality in Communicative Events, *American Anthropologist*, 81, 773-790.
- Iten, C. (1998) The meaning of although: a relevance theoretic account. UCL Working Papers in *Linguistics*, 10, 81-108.
- Jiang, N., & Nekrasova, T. M. (2007). The Processing of Formulaic Sequences by Second Language Speakers. *Modern Language Journal*, 91(3), 433-445.
- Kasper, G. (1996). Introduction: interlanguage pragmatics in SLA, in *Studies of Second Language Acquisition*, 18, 145-148.
- Kasper, G. (2007). Pragmatics in second language learning: An update. In A. Valdman (Ed.), *Language learning celebrates 30 years of AAAL*
<http://www.aaal.org/index.php?id=53>
- Kasper, G., & Blum-Kulka S. (1993). *Interlanguage Pragmatics*, New York: Oxford University Press.
- Kasper, G., & Dahl, M., (1991). Research Methods in Interlanguage Pragmatics, in *Studies in Second Language Acquisition*, 13, 213-247.
- Kasper, G. & Rose, K.R. (2002). *Pragmatic Development in a Second Language*. Malden, Mass. etc.: Blackwell.

- Kasper, G. & Roever, C. (2005). Pragmatics in second language learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 317-334). Mahwah, NJ: Erlbaum.
- Kasper, G., & Rose, K. R. (2002). *Pragmatic development in a second language*. Oxford: Blackwell.
- Kilgarrieff, A. (1995). BNC database and word frequency lists. Available at <http://www.kilgarrieff.co.uk/bnc-readme.html> (accessed December, 2009)
- Kilgarrieff, A. 2009. "Simple maths for keywords". In *Proc. Corpus Linguistics*, Liverpool, UK.
- Kilgarrieff, A., Rychly, P., Smrz, P. & Tugwell, D. (2004). The Sketch Engine, *Proc EURALEX 2004* (pp. 105-116), Lorient, France. <http://ca.sketchengine.co.uk/open/>
- Klimt, B. & Yang, Y. 2004. "Introducing the Enron corpus". In *Proceedings of First Conference on Email and Anti-Spam (CEAS)*, July 30-31, Mountain View, CA, USA. Available at <http://www.ceas.cc/papers-2004/>
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Lampert, A., Dale, D., & Paris, C. (2008). Requests and commitments in email are more complex than you think: Eight reasons to be cautious. In David Powers & Nicola Stokes (Eds.), *Proceedings of Australasian Language Technology Workshop* (pp. 55-63). Hobart, December 8-10.
- Lee, C. (2010). An exploratory study of interlanguage pragmatic comprehension of young learners of English. *Pragmatics*, 20, 3, 343-373.

- Lee, D. (2001). Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology*, 5, 37-72.
- Leech, G.N. (1983). *Principles of Pragmatics*. London: Longman.
- Leech, G., Deuchar, M., Hoogenraad, R. (1982). *English Grammar for Today. A new introduction*, London: Macmillan.
- Lenk, U. (1998). Discourse markers and global coherence in conversation, *Journal of Pragmatics*, 30, 245-257.
- Lexile Framework for Reading. (2008). *Lexile Analyser*. Retrieved Sep 30, 2010, from <http://www.lexile.com>
- Lexile (2008). *Lexiles: A System for Measuring Reader Ability and Text Difficulty: A Guide for Educators*. Retrieved Sep 30, 2010, from http://teacher.scholastic.com/products/sri_reading_assessment/pdfs/SRI_ProfPaper_Lexiles.pdf
- Levinson, S. (1983): *Pragmatics*. Cambridge: Cambridge University Press.
- Li, D. (2000). The pragmatics of making requests in the L2 workplace: A case study of language socialization. *Canadian Modern Language Review*, 57 (1), 58-87
- Liu. A.L.E. (2002). *A Corpus-based Lexical Semantic Investigation of Verb- Noun Miscollocations in Taiwan Learners' English*, unpublished MA thesis, Tamkang University, Taiwan.
- Liu. A.L.E. (2009). Discourse markers in two registers: an experiment on how L2 learners link coherence relations in formal and informal settings. Unpublished manuscript.

- Liu, A.L.E. (2010a). Multi-word discourse markers in use: the patterns in spoken and written corpora, *ICAME 31*, Giessen University, Germany.
- Liu, A.L.E. (2010b). Use of multi-word discourse markers in the EFL writing: evidence from the Bulgarian, Chinese, Japanese and Russian subcorpora of the International Corpus of Learner English, *The 5th IVACS Conference*, June, University of Edinburgh, United Kingdom.
- Lorenzo-Dus, N. (2001). Compliment responses among British and Spanish university students: A contrastive study. *Journal of Pragmatics*, 33, 107-127.
- Mann, W.C. & Thompson, S.A. (1986). Relational Propositions in Discourse, *Discourse Processes*, 9, 57-90.
- Manning, C. & Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*, MIT Press. Cambridge, MA.
- Mauranen, A. (1993). *Cultural Difference in Academic Rhetoric: A Textlinguistic Study*. Frankfurt: Peter Lang.
- McCallum, A., Wang, X., & Corrada-Emmanuel, A. (2007). Topic and role discovery in social networks with experiments on enron and academic email. *Journal of Artificial Intelligence Research*, 30, 249-272.
- McCarthy, M. (1991). *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.
- McCarthy, M. & R. Carter (1994). *Language as Discourse: Perspectives for Language Teaching*. London: Longman.
- McEnery, T. & Xiao, R. (2004). Swearing in modern British English: the case of FUCK in the BNC. *Language and Literature*, 13 (3), 235-268.

- McEnery, T., Xiao, R. Tono, Y. (2006). *Corpus-based Language Studies. An Advanced Resource Book*. Routledge, London.
- Meyerhoff, M. (1998). Accommodating your data: The use and misuse of accommodation theory in sociolinguistics. *Language and Communication*, 18, 205-225.
- Milton, J.C.P., & Tsang, E. S-C. (1993). A corpus-based study of logical connectors in EFL students' writing : directions for future research. *Studies in lexis*. Proceedings of a seminar on lexis organized by the Language Centre of the HKUST, Hong Kong , 6-7 July, pp. 215-246.
- Müller, S. (2005). *Discourse Markers in Native and Non-Native English Discourse*. Amsterdam: John Benjamins.
- Myles, F., Hooper, J.& Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48(3): 323-363.
- Narita, M., Sato, C., & Sugiura, M. (2004). Connector usage in the English essay writing of Japanese EFL learners. Paper presented at the *Fourth International Conference on Language Resources and Evaluation*.
- Narita, M. & Sugiura, M. (2006) 'The use of adverbial connectors in argumentative essays by Japanese EFL college students'. *English Corpus Studies* 13, 23-42.
- Neff, J., Ballesteros, F., Dafouz, E., Martínez, F. & Rica, J.P. (2007). A contrastive functional analysis of errors in Spanish EFL university writers' argumentative texts: corpus-based study, in E. Fitzpatrick (ed.) *Corpus Linguistics beyond the Word: Corpus Research from Phrase to Discourse* (pp. 203-225), Amsterdam: Rodopi.

- Nesselhauf N. (2005). *Collocations in a Learner Corpus*. Amsterdam: Benjamins.
- New, B., Ferrand, L., Pallier, C., and Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: new evidence from the English Lexicon Project. *Psychon Bull Rev.*, 13 (1), 45-52.
- Nikula, T. (1996). *Pragmatic Force Modifiers: A Study in Interlanguage Pragmatics*, University of Jyväskylä, Jyväskylä, Finland.
- O'Keeffe, A., Clancy, B. & Adolphs, S. (2011). *Introducing pragmatics in use*. London: Routledge
- Olshtain, E., & Blum-Kulka, S. (1985). Degree of approximation: Nonnative reactions to native speech act behaviour. In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 303-325). Rowley, MA: Newbury House.
- Pawley, A. & Syder, F. (1983). Two Puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. Richards and R. Schmidt (Eds.), *Language and Communication*. Longman.
- Peterson, K., Hohensee, M., & Xia, F. (2011). Email formality in the work place: a case study on the enron corpus. In *Proceedings of 49th Annual Meeting of the Association for Computational Linguistics (ACL '11)*.
- Pulcini, V. and C. Furiassi, (2004). Spoken interaction and discourse markers in a corpus of learner English. In Partington A., J. Morley and L. Haarman (eds) *Corpora and Discourse*. Bern: Peter Lang. 107-123.
- Quirk, R., Greenbaum, S., Leech G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

- Redeker, G. (1991). Review article: Linguistic markers of discourse structure. *Linguistics*, 29(6): pp.1139-72.
- Reed, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Reid, J. (1993). *Teaching ESL Writing*. Englewood Cliffs, N.J.: Prentice-Hall.
- Reid, T. B.W. (1956). Linguistics, Structuralism, and Philology, *Archivum Linguisticum*, 8, 28-37.
- Richard, W. and Joseph, T. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101, 2, 192-212.
- Richards, J. C., Schmidt, R., Kendricks, H. & Kim, Y. (2002). *Longman dictionary of language teaching and applied linguistics* (3rd ed.). Harlow, Essex: Longman.
- Rieber, S. (1997). Conventional implicatures as tacit performatives. *Linguistics and Philosophy*, 20 (1), 51-72.
- Robinson, M. (1992). Introspective methodology in interlanguage pragmatics research. In G. Kasper (Ed.), *Pragmatics of Japanese as native and target language* (Technical Report #3, pp. 27-82). Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.
- Romero Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English, *Journal of Pragmatics* 34, 769-784.
- Rose, K.R. (2000). An exploratory cross-sectional study of interlanguage pragmatic development, *Studies in Second Language Acquisition*, 22, 27-67.
- Rose, K. R. & Ng, K-f. C. (2001). Inductive and deductive teaching of compliments and compliment responses. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 145-170). New York: Cambridge University Press.

- Rubin, J. (1968). Language and education in Paraguay. In J. A. Fishman, C. A. Ferguson, & J. Das Gupta (Eds.), *Language problems in developing nations*. New York: Wiley.
- Rudolph, E. (1996). *Contrast. Adversative and concessive relations and their expressions in English, German, Spanish, Portuguese on Sentence and Text Level*. Berlin/New York: de Gruyter
- Sanders, T.J.M., Spooren, W.P.M. & Noordman, L.G.M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, 15, 1-35.
- Schachter, J. (1990). Communicative competence revisited. In B. Harley, P. Alien, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency*. New York: Cambridge University Press.
- Schauer, G. A. (2006). Pragmatic Awareness in ESL and EFL Contexts: Contrast and Development. *Language Learning*, 56, pp269–318.
- Schiffrin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics, In G. Kasper and S. Blum-Kulka (Eds.), *Interlanguage Pragmatics* (pp. 21-43), New York: Oxford University Press.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on attention and awareness in learning. In R. Schmidt, Ed., *Attention and awareness in foreign language learning* (pp. 1-63). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.
- Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*, Philadelphia: John Benjamins.
- Schmitt, N. & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1-22). Philadelphia: John Benjamins.
- Schmitt, N., Grandage, S. & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 127-151). Philadelphia : John Benjamins.
- Schourup, L. (1999). Discourse markers: Tutorial review, *Lingua*, 107, 227-265.
- Scollon, R. & Scollon, S. W. (2001). *Intercultural Communication*, 2nd ed. Blackwell, Oxford.
- Searle, J. R. (1969). *Speech Acts. An Essay in the Philosophy of Language*. Cambridge
- Shaw, P. & Liu, E. (1998). What develops in the development of second language writing?, *Applied Linguistics*, 19 (2), 225-54.
- Siepmann, D. (2005). Discourse Markers across Languages. A contrastive study of second-level discourse markers in native and non-native text with implications for general and pedagogic lexicography. Abingdon/New York: Routledge.
- Simpson, R. C., S. L. Briggs, J. Ovens, and J. M. Swales. (2002) *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.
- Sinclair, J. & Coulthard, M. (1975). *Towards an Analysis of Discourse*. London: Oxford UP.

- Sperber, D. and Wilson, D. (1986). *Relevance: communication and cognition*. Oxford: Blackwell. (Second edition 1995.)
- Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2007). *The Lexile framework for reading* (Technical report). Durham, NC: Metametrics.
- Street, R. L. and Giles, H. (1982). Speech accommodation theory: A social cognitive approach to language and speech behaviour. In M. Roloff and C. R. Berger (eds.), *Social cognition and communication* (pp. 193-226). Beverly Hills, CA: Sage.
- Stubbe, M. & Holmes, J. (1995). You know, eh and other 'exasperating expressions': an analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language and Communication*, 15, 1, 63-88.
- Stubbs, M. (1980) *Language and Literacy: The Sociolinguistics of Reading and Writing*. London: Routledge & Kegan Paul.
- Stubbs, M. (1983). *Discourse Analysis*. Chicago, IL: The University of Chicago Press.
- Styler, W. 2011. "The EnronSent Corpus". *Technical Report 01-2011*, University of Colorado at Boulder Institute of Cognitive Science, Boulder, CO.
- Swales, J.M. (1990). *Genre analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Swales, J. & Feak, C. (2004). *Academic Writing for Graduate Students: Essential tasks and skills, Second Edition*, Ann Arbor: University of Michigan Press.
- Taguchi, N. (1999). L2 proficiency influence on learners' pragmatic realization: A speech act of request making . Paper presented at the *AILA International Conference*, Tokyo, Japan.

- Taguchi, N. (2002). An application of relevance theory to the analysis of L2 interpretation processes: The comprehension of indirect replies. *International Review of Applied Linguistics*, 40, 151–176.
- Taguchi, N. (2005). Comprehending implied meaning in English as a second language. *Modern Language Journal*, 89, 543–562.
- Taguchi, N. (2011). The Effect of L2 Proficiency and Study-Abroad Experience on Pragmatic Comprehension. *Language Learning*, 61: no. doi:10.1111/j.1467-9922.2011.00633.x
- Tang, E. & Ng, C. (1995). A Study on the Use of Connectives in ESL Students' Writing. *Perspectives Working Papers*, 5 (1), 38-45. Hong Kong.
- Tankó, G. (2004). The use of adverbial connectors in Hungarian university students' argumentative essays. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 157–81). Amsterdam: John Benjamins.
- Trosborg, A. (1995). *Interlanguage Pragmatics: Requests, Complaints and Apologies*, Mouton de Gruyter, New York.
- Trudgill, P. (1983). *Sociolinguistics: an introduction to language and society*. London: Penguin Books.
- Urgelles-Coll, M. (2010). *Syntax and Semantics of Discourse Markers*. London: Continuum
- van Dijk, T. A. (1977). *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. London: Longman.
- Van Dijk, T. (1979). Pragmatic connectives, *Journal of Pragmatics*, 3, 447-56.

- Vande Kopple, W. (1985). Some Exploratory Discourse on Metadiscourse, *College Composition and Communication*, 36, 82-93.
- Waibel, B. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students*, Freiburg: Albert-Ludwigs-Universität Freiburg.
- Warner, R.G. (1985). *Discourse Connectives in English*, New York: Garland.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: a review. *Applied Linguistics*, 16 (2), 180-205.
- Wikborg, E. & Björk, L. (1989). *Sammanhang i text. En empirisk undersökning och skrivpedagogiska konsekvenser. [Coherence in text. An empirical study with pedagogical consequences for composition teaching.]* Uppsala: Hallgren and Fallgren.
- Williamson, G. L. (2008). A Text Readability Continuum for Postsecondary Readiness. *Journal of Advanced Academics*, vol. 19 (4), 602-632.
- Wilson, D. & Sperber, D. (1993). Linguistic form and relevance. *Lingua*, 90, 1/2: 1-25.
- Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal*, 20, 1-15.
- Wray, A. (1999). Formulaic Language in learners and native speakers, *Language Teaching*, 32, 213-231.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge : Cambridge University Press.
- Wray, A. (2004). Formulaic language learning on television. Formulaic sequences. Acquisition, processing and use, In N. Schmitt (Ed.), *Language Learning and Language Teaching* 9 (pp.249-268), Amsterdam: John Benjamins.

Yule, G. U. (1912). On the methods of measuring association between two attributes.

Journal of the Royal Statistical Society, 75 (6), 579-652.

Zamel, V. (1983). Teaching those missing links in writing. *ELT Journal*, 37, 22-29.

Zelinsky, G. J. and Murphy, G.L. (2000). Synchronizing visual and language processing:

an effect of object name length on eye movements. *Psychological Science*, 11: 125.

Appendix 1. List of the target MDMs

Additive

what's more, not only that, in addition, in addition to this, in addition to that, on top of that, on top of this, besides that, besides this, add to this, add to that, and also

Appositive

that is to say, that is, in other words, what I'm saying is, what this means is, what that means is, what it means is

Causal

as a result, as a result of this, as a result of that, because of that, because of this, as a consequence, for that reason, for this reason

Contrastive

on the other hand, on the contrary, in contrast, by contrast

Concessive

despite that, despite this, having said that, that said, that being said, having said this, this said, this being said, but then again, then again, but again, in spite of that, in spite of this, at the same time, on the other hand, but still

Appendix 2. Study 1: the nine passages for the six target MDMs and the neutral MDMs

The passages below are shown as they were presented as slides to participants. Each slide has one or two sentences. The slide number is shown next to the sentence(s). For example, S1 means Slide 1. Item 1 to 6 are the six passages for the target MDMs. Item 7 to 9 are passages for the neutral MDMs.

1. Informal Additive MDM: *what's more*

S1. Well I was just gonna say that. when I was at the Notting Hill Carnival yesterday

S2. it was like people were seriously bringing their push-chairs and little children into a crowd.

S3. It was actually a a massive crowd and you would get you know kind of.

S4. I don't know it was really awful. Yeah. I couldn't believe they had children.

S5. They took children with them. Or they took a dog or something. Erm.

S6. What's more, not just that it's doing the kid's eardrums damage but also they'll use the child as a battering ram.

S7. Yes. Smacking this thing into you legs to get through. As if because they've been they've got kids you move.

S8. Please press the spacebar when ready for the T/F statement.

S9. This text is about how some people brought their books to a carnival. (1=True, 2=False)

S10. Press "ENTER" to continue.

2. Formal Additive MDM: *in addition*

- S1. This supports the idea that the sample of left-handers used would have had considerable time to learn to modify their motor skills appropriately
- S2. to adapt satisfactorily to their environment despite being disadvantaged when faced with right-handed tools.
- S3. The findings also demonstrated that left-handers were no more likely than right-handers to be significantly 'right brained'.
- S4. The reason that left-handers did not show more evidence of right hemisphere attributes may be because they are constantly encouraged to adopt left hemisphere characteristics.
- S5. In schools, emphasis is placed on objective learning. The ability to excel in subjects such as mathematics and science are well rewarded.
- S6. In addition, the world is geared towards logical thinking.
- S7. Only in recent times has creativity and lateral thinking been truly encouraged at basic education level.
- S8. Press "ENTER" for the T/F statement.
- S9. This passage is about how left-handers are better in chemistry. ("YES"=True, "NO"=False)
- S10. Press "ENTER" to continue.

3. Informal Causal MDM: *because of that*

- S1. One of her we got Did I tell you we're getting invited to Olga's wedding?

S2. Well it was all off. Mm. Er er apparently she took er a more serious look at him and thought Oh no.

S3. I think he liked too much T V and things like that that she wasn't interested in.

S4. She's got all her marbles hasn't she Olga. She's got I mean everything was booked.

S5. it I mean it er was it er two weeks before the actual date that it was all off.

S6. Because of that, we'd just sent a cheque. You know

S7. I mean by the time we'd faffed about with lists whether or not people were going to return it and all that.

S8. Press "ENTER" for the T/F statement.

S9. This passage is about how a girl cancelled her wedding all in a sudden.

("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

4. Formal Caudal MDM: in consequence

S1. There was a gross overspend that year, which led, in part, to the replacement of Francis Pym by John Nott, as will be recounted in chapter 8.

S2. Hitherto departments had not been allowed to carry forward underspends into the next financial year

S3. The Treasury was allowed to penalize departments by deducting overspends from the next year's targets.

S4. In consequence, there was always a scramble to spend in the last half of the financial year when the danger of an underspend was becoming apparent.

S5. It was usually too late to accelerate existing contracts or to initiate new ones

S6. The 'Bow-Wave Phenomenon' turned into the 'Barrack Square Syndrome', in which the only way to mop up money

S7. was to spend it on a few quickly and easily let contracts, like refurbishing barrack squares and military roads.

S8. Please press 'ENTER' for the T/F statement.

S9. This passage talks about ways of spending the remaining budget before the end of a financial year. (1=True, 2=False)

S10. Press "ENTER" to continue.

5. Informal Concessive MDM: *but then again*

S1. I used to tinker about with cars sort of like when I was sixteen to twenty two twenty three

S2. My granddad was always there because he loved pulling engines out of cars

S3. You know he was more Er I know it's probably a horrible thing to say

S4. he was probably more of a father in that respect than my own father was

S5. dad hadn't got the time because he was working supporting seven seven of us you know.

S6. I did a lot more with my granddad because that was his my interest as well as his own you know which hit you know hard.

S7. But then again, life isn't a bed of roses is it.

S8. Press "ENTER" for the T/F statement.

S9. The passage is about how a grandmother influences a granddaughter.

("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

6. Formal Concessive MDM: *despite that*

S1. There has been much debate over whether what we are seeing is true imitation and not just some form of conditioning.

S2. In previous research, examiners have been found to alter their rhythm of tongue protrusions as a function of the infant's response, thus encouraging the possibility for conditioned imitation.

S3. To rule out this factor, they ensured that examiners were unaware that they were testing imitation and blind judges were used to label the newborns' expressions.

S4. Each newborn was shown four gestures in a random order for 15 seconds.

S5. It was found that 12-21 day old babies imitated both facial and manual gestures during a 20 second response period.

S6. Despite that, some research has proposed that imitation is merely an innate releasing mechanism

S7. where these response gestures are just fixed action patterns released by a corresponding adult gesture.

S8. Press "ENTER" for the T/F statement.

S9. This passage is about how newborn puppies respond and imitate adult behaviours. ("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

7. Neutral Additive MDM: *besides that*

S1. In the case of an exchange error found in LOWING THE MORN (mowing the lawn), the L and M was mistakenly copied,

S2. while a repetition error in BEEF NEEDLE (beef noodle), the EE was not deleted after being copied.

S3. The example A BURLY BIRD (An early bird) presented an anticipatory error where the phoneme /b/ of bird was added to the beginning of early.

S4. Note that the indefinite article was also changed to conform to the grammatical rule.

S5. This suggested the existence of a 'monitoring device' that prevents a compounding of errors.

S6. Besides that, it also reveals some information about stages of speech production--phonemic disorder has to occur before the indefinite article is given its phonological form,

S7. or the morphological rule have to reapply after the initial error has occurred.

S8. Press "ENTER" for the T/F statement.

S9. The passage is about sound displacements in speech. ("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

8. Neutral Causal MDM: *for that reason*

S1. Joyce uses a similar technique to develop a character in 'The Sisters,' the first story in the collection.

S2. It is written from the point of view of a young boy, and Joyce is careful, to keep the language and opinions of the piece consistent with his protagonist.

S3. For that reason, he changed the following passage which was written in an adult cadence:

S4. The ceremonious candles in the light of which the Christian must take his last sleep.'

S5. The sentence was replaced with the much more straightforward and child-like:

S6. 'The reflection of candles on the darkened blind for I knew that two candles must be set at the head of a corpse.'

S7. By simplifying the language and extending the sentence length, Joyce ensures that it is much more in keeping with a younger narrator.

S8. Press "ENTER" for the T/F statement.

S9. This passage is about how a writer develops his protagonist by modifying the language used. ("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

9. Neutral Concessive MDM: *that being said*

S1: As has been touched upon, it is arguably the exclusive male elite who have reaped the rewards of the new capitalist economy.

S2: It is this group alone which has set the agenda and which heads the process of global change.

S3: The principle model which can be been applied to most countries, seems to place men in the productive public sector and the concentration of unpaid female labour in the reproductive sector.

S4: The patriarchal structure of the male as breadwinner and the female caregiver still stands true.

S5: That being said, within this patriarchal process we had witnessed a shift to an 'open economy free market approach'.

S6: Within this shift, Standing refers to the importance of the 'renewed surge of feminization of labour activity'.

S7: This surge has had mixed effects on both ' first ' and ' third ' world women.

S8: Press "ENTER" for the T/F statement.

S9: This passage is about the change of women's roles in schools. ("YES"=True, "NO"=False)

S10: Press "ENTER" to continue.

Appendix 3. Study 1: Experimental List 1

This experimental list includes three passages from the targeted semantic categories: Additive, Causal, and Concessive, and 12 filler passages.

The passages below are shown as they were presented as slides to participants. Each slide has one or two sentences. The slide number is shown next to the sentence(s). For example, S1 means slide 1. Item 1 to 3 are the passages for the target MDMs. Item 4 to 15 are fill passages.

Item 1. *what's more* in the informal context

S1. Well I was just gonna say that. when I was at the Notting Hill Carnival yesterday

S2. it was like people were seriously bringing their push-chairs and little children
into a crowd.

S3. It was actually a a massive crowd and you would get you know kind of.

S4. I don't know it was really awful. Yeah. I couldn't believe they had children.

S5. They took children with them. Or they took a dog or something. Erm.

S 6: What's more, not just that it's doing the kid's eardrums damage but also they'll
use the child as a battering ram.

S7. Yes. Smacking this thing into you legs to get through. As if because they've been
they've got kids you move.

S8. Please press the spacebar when ready for the T/F statement.

S9. This text is about how some people brought their books to a carnival. (1=True, 2=False)

S10. Press "ENTER" to continue.

Item 2. *in consequence* in the formal context

S1. There was a gross overspend that year, which led, in part, to the replacement of Francis Pym by John Nott, as will be recounted in chapter 8.

S2. Hitherto departments had not been allowed to carry forward underspends into the next financial year

S3. The Treasury was allowed to penalize departments by deducting overspends from the next year's targets.

S4. In consequence, there was always a scramble to spend in the last half of the financial year when the danger of an underspend was becoming apparent.

S5. It was usually too late to accelerate existing contracts or to initiate new ones

S6. The 'Bow-Wave Phenomenon' turned into the 'Barrack Square Syndrome', in which the only way to mop up money

S7. was to spend it on a few quickly and easily let contracts, like refurbishing barrack squares and military roads.

S8. Please press 'ENTER' for the T/F statement.

S9. This passage talks about ways of spending the remaining budget before the end of a financial year. (1=True, 2=False)

S10. Press "ENTER" to continue.

Item 3. *that being said* in the neutral context

- S1. As has been touched upon, it is arguably the exclusive male elite who have reaped the rewards of the new capitalist economy.
- S2. It is this group alone which has set the agenda and which heads the process of global change.
- S3. The principle model which can be been applied to most countries, seems to place men in the productive public sector and the concentration of unpaid female labour in the reproductive sector.
- S4. The patriarchal structure of the male as breadwinner and the female caregiver still stands true.
- S5. That being said, within this patriarchal process we had witnessed a shift to an 'open economy free market approach'.
- S6. Within this shift, Standing refers to the importance of the 'renewed surge of feminization of labour activity'.
- S7. This surge has had mixed effects on both ' first ' and ' third ' world women.
- S8. Press "ENTER" for the T/F statement.
- S9. This passage is about the change of women's roles in schools. ("YES"=True, "NO"=False)
- S10. Press "ENTER" to continue.

Fillers

Item 4

- S1. The hostessing employment world is dominated by women not men which is key as it demonstrates that
- S2. generally in Japanese society it is believed that there is 'one gender of servicer-female and one gender of serviced-male.
- S3. Allison continues to demonstrate this concept in relation to home life and marriage.
- S4. The word for housewife, 'oku' literally means 'inside' which reflects their job of managing all things within the house
- S5. which includes the 'management of all matters of home, family and domestic maintenance.
- S6. By contrast, the single constant expectation of a father/husband is to be the financial supporter of the house.
- S7. Here Allison illustrates again the idea of women servicing and doing everything for men- 'she is the slave, he the master'.
- S8. Press "ENTER" for the T/F statement.
- S9. This passage is about roles of woman and man in the English society.
("YES"=True, "NO"=False)
- S10. Press "ENTER" to continue.

Item 5

- S1. It is theorised that the parental inexperience and anxiety associated with overprotective parents instils greater anxiety in children.

- S2. Children with overprotective fathers have a tendency towards being anxious, reserved and subsequently introverted.
- S3. Children who are overprotected also lack the opportunities to learn social skills which may result in a more sociable and extroverted personality.
- S4. Putallaz reveals that children with authoritarian parents have poorer peer relations; the reverse is true for motherly overprotection.
- S5. In other words, the more overprotective mothers are, the more extroverted their children tend to be.
- S6. It could be that children are able to rebel more successfully against their mothers.
- S7. The more authoritarian mothers are, the more inclined their children are to rebel.
- S8. Press "ENTER" for the T/F statement.
- S9. This passage is about how parents influence their children's personality.
("YES"=True, "NO"=False)
- S10. Press "ENTER" to continue.

Item 6

- S1. On almost every metric, Tesco outperforms Somerfield. On some metrics, Somerfield is in a particularly precarious position.
- S2. Of special concern are the liquidity and solvency ratios, given our position as a supplier.

- S3. Somerfield pays consistently later than Tesco and that, coupled with the knowledge of the overall stability of the firm, gives cause for concern.
- S4. More information would be useful in order to more accurately assess Somerfield and Tesco: full report and accounts,
- S5. including more detail on capital expenditure in particular for Somerfield Company structure Management team Strategy and goals of each company.
- S6. In summary, our company should reconsider the reliance it places on Somerfield and attempt to find other purchasers for our products.
- S7. Not to do so represents a huge risk to our future.
8. Press "ENTER" for the T/F statement.
- S9. This passage is about assessing Tesco and Sainsbury's. ("YES"=True, "NO"=False)
- S10. Press "ENTER" to continue.

Item 7

- S1. Maternalism is the idea that motherhood is a natural desire in all women. This is evident in many arguments supporting motherhood.
- S2. Motherhood provides women with a sense of identity, meaning and status in both the society and the family.
- S3. I will now take each aspect of motherhood and illustrate how motherhood can be seen as a woman's destiny.
- S4. To begin with, motherhood provides women with a form of feminine identity and status.

- S5. It is through motherhood in which women become socially recognised as a 'real' woman, who has fulfilled her destiny in life.
- S6. A woman never achieves full status until she has children.
- S7. There is an implicit assumption that parenthood is intrinsic to adult female identity; to not have children would be seen as failing the ideology of womanhood that society perpetuates.
- S8. Press "ENTER" for the T/F statement.
- S9. This passage is about how becoming a parent shapes the identity of a woman.
("YES"=True, "NO"=False)
- S10. Press "ENTER" to continue.

Item 8

- S1. I believe this is primarily due to the stress caused by the nature of dementia. The patient is rarely able to give adequate feedback.
- S2. And although the patient will require constant care and attention, they may not even recognise their caregiver.
- S3. Understandably, this could be extremely stressful for the caregiver, particularly if the patient were a spouse or a parent,
- S4. as the situation would be deeply upsetting and could be the cause of major psychological effects.
- S5. In fact, caring for a dementia patient results in a higher risk of depression, suicide and other health related problems.

S6. Both the emotional and physical strain placed upon the caregiver and the patient is not helped by the fact

S7. that many caregivers are reluctant to ask for help, possibly due to feeling guilty about being an inadequate caregiver.

S8. Press "ENTER" for the T/F statement.

S9. This passage is about the stress experienced by school teachers. ("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

Item 9

S1. This essay aims to provide a review of ageing literature on selective attention, discussing the experiments and theories that arise from them,

S2. illustrating that our knowledge from this type of research cannot only provide valuable insights into other theories of ageing but can also help us to improve life in old age.

S3. The most commonly used method is the visual search task, which is where target items and distractors are presented in the same display.

S4. Traditional research found that when participants were presented with a cue indicating where a target item would subsequently be presented,

S5. only younger adults showed decreased reaction times.

S6. That is to say, older adults did not use the cues to facilitate their search even though they attended to them,

S7. as invalid cues were found to increase their reaction times.

S8. Press "ENTER" for the T/F statement.

S9. This passage is about how ageing affects older adults' attention in a visual search task. ("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

Item 10

S1. Although she doesn't much like the wrestling she'd rather be same place; i am rather than somewhere completely different

S2. If i go wrestling then Jill will go wrestling too. if i go to the opera Jill will also go to the opera

S3. She likes going to the opera but she'll go there just to meet me how nice.

S4. In that case, Jack knowing Jill 's preferences knows that whatever he does it will pay Jill once she knows what he's done, to do the same.

S5. He can then use this information to his own advantage because by moving first

S6. he can say aha i can go wrestling because i know that even though Jill would rather go to the opera

S7. if she knows i am going to the wrestling she 'll go to the wrestling too.

S8. Press "ENTER" for the T/F statement.

S9. This passage is about how a girl is crazy about wrestling. ("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

Item 11

S1. That made me feel even worse. Mm. I was yes. Cos the the flu sort of it sort of went and then it came back.

S2. It was that kind of a flu. You One day I think ooh I feel much better I'm all right

S3. couple of days after it was back again you know.

S4. I've still got a bit of it now and I I had to go and get some more antibiotics this morning cos like oh I just feel.

S5. My tongue feel heavy and my head feels weird.

S6. I've got this horrible headache keep coming and going. Mm.

S7. All in all, em to me anyway it was not Christmas like I know. That was it.

S8. Press "ENTER" for the T/F statement.

S9. This passage is about a person who had a flu during New Year's time.

("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

Item 12

S1. We actually had a bomb alert Monday. And the kids were there.

S2. Er I was up in the canteen and er you've never seen a canteen I mean I've I've never been in the canteen before.

S3. when that's gone off and people crashing into each other because in eating your meal you run. You go.

S4. Two people actually left the store during that alert.

S5. For one thing, you can't even go to the toilet while that alert is on till the all clear. You stay on your department.

- S6. People were walking around with these new new members of staff.
- S7. By rights we shouldn't I don't think that we should actually be checking the goods in the store for these bombs.
- S8. Press "ENTER" for the T/F statement.
- S9. This passage is about what happened when a bomb alert went off. ("YES"=True, "NO"=False)
- S10. Press "ENTER" to continue.

Item 13

- S1. In fifteen years time two thousand quid won't be worth postage stamps sort of thing
- S2. Leave it till then and then give us a postage stamp. There's.
- S3. The whole point is that erm as the family unit I think we've we have got to these days
- S4. to to hang together er or we'll hang separately. Well.
- S5. As a matter of fact, we didn't but what we did what we did do is Grandma and Grandpa every Wednesday when the shop shut for the half day.
- S6. they would come down and they'd brought er er a joint of meat with them.
- S7. They would bring a joint of meat and often very a bottle of wine and this was our little injection of luxury.
- S8. Press "ENTER" for the T/F statement.
- S9. This passage is about how he and his grandma and grandpa used to get together to enjoy some wine and food. ("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

Item 14

S1. Erm I kept going in to look at this you know this book.

S2. it was quite expensive so I was kind of looking for a discount.

S3. we kind of got talking and and I never would ask him for discount

S4. As we got to know each other better and better there was less and less chance of
me trying to give it a try you know.

S5. When Christmas came around and he bought it for me for a Christmas present.

Which is really sweet. Just in the shop.

S6. In the meantime, I had been out and bought it myself from somebody else when I
know he knew he was on his lunch break.

S7. Yeah. Cos I really wanted it but I wouldn't ask him for it.

S8. Press "ENTER" for the T/F statement.

S9. This passage is about how to choose a car as a Christmas present. ("YES"=True,
"NO"=False)

S10. Press "ENTER" to continue.

Item 15

S1. I wouldn't go through the education system now.

S2. They I mean you've stopped they they've stopped the erm maintenance grant
haven't they.

S3. They they froze that didn't they. Reduced it by ten per cent each year. Erm they're stopping tuition fees as well aren't they.

S4. I mean it's it's not it's erm you know it it it in fact really it it it even erm it it causes

S5. I think that kind of thing causes more polarization between the classes if you like

S6. the wealthy people they they'd be fine they can still afford to send their their kids to school. Then to university because erm

S7. In any case, they had to pay the maintenance class and tuition fees anyway didn't they.

S8. Press "ENTER" for the T/F statement.

S9. This passage is about how the cut of grant affects students. ("YES"=True, "NO"=False)

S10. Press "ENTER" to continue.

Appendix 4. Study 2: Multiple-choice Discourse Completion Task

Thank you for helping out with this study!

In what follows you will find 10 short passages (excerpts), which are taken directly from several language databases without any modification. This means that some of them might include spelling mistakes and ungrammatical parts. These, however, will not affect the understanding of the passage.

Each passage has one short phrase missing. What you have to do is to read the passages and choose the most appropriate missing phrase from the 10 choices provided. Note that some of the phrases in the 10 choices might look similar in meaning. Please read the passage carefully and choose the most appropriate one. Then please answer the very simple True/False question below each passage.

Before we start, below is an example.

(1) First of all (2) But then again(3) As a result(4) What I'm saying is(5) To begin with
(6) Because of that (7) That is (8) What's more(9) Despite this/that(10) In addition

...If the unfavourable treatment included a significant sexual element to which a person of the opposite sex would not have been vulnerable, that amounted to discrimination. More than a decade after the introduction of legislation against sex or race discrimination, it remains true that only a minority of senior business executives are women or members of racial minority groups. There are no doubt many reasons for this; one of them is that unspoken, perhaps sometimes subconscious, prejudice remains deeply ingrained in many management structures. _____, surprisingly few complaints about discrimination are made to industrial tribunals each year. Even fewer succeed and the average level of compensation in those cases is just a few hundred pounds.

Q: This passage is about discrimination toward male executives. (T or F?)

Once you have read it, you will see that the phrase 'Despite this/that' is the most appropriate phrase to put in the blank because it sounds natural. So, number 9 will be put in the passage. And as the passage is not about the discrimination toward male executives, the answer for the T/F question is F.

Now, the ten phrases below are choices for the following 10 passages (from p.2-5). Please take some time to familiarize yourself with the following phrases.

(1) Despite this/that (2) What's more (3) As a result (4) That is (5) In addition
(6) But then again (7) Because of that (8) What I'm saying is (9) First of all(10) To begin with

These ten passages will again appear on top of each page. NOTE that the sequence orders of the 10 phrases are DIFFERENT on every page. For example, when you are working on the passages on page 3, please refer to the choices provided on top of page 3 only. This is an UNTIMED task so please take your time in understanding the passages and in choosing. If you are ready, please turn to the next page to start.

(1) In addition (2) But then again (3) First of all (4) As a result (5) Despite this/that (6) That is (7) What's more (8) To begin with (9) What I'm saying is (10) Because of that

1. ...Although there was first a certain amount of debate about what the nature and title of the broadcasts should be, Lewis began to do this in the late summer of 1941, taking the train from Oxford to London every Wednesday evening, and broadcasting from 7.45 to 8.00 p.m. Sound-broadcasting is a particular skill, not necessarily related to literary ability though impossible without it. _____, one needs the literate ability to express oneself clearly; but one also needs the right voice and the ability to be concise. Lewis's broadcasts during the war were in three series, and they were written up (published more or less as spoken over the air) as Broadcast Talks (1942), Christian Behaviour (1943) and Beyond Personality (1944).

Q: This passage is about the broadcasting someone did in 2010. _____ (T or F?)

2. ...I would like to weigh in on this, if you don't mind. _____, I requested meetings with the audit and EHS folks to understand our procedures and our status. So, no fair ranting at Don. Secondly, your note brings to mind what I find the most troublesome about our process as it currently stands. It is not now, nor will it be in the future the audit department's responsibility to ensure that corrective actions are taken as a result of the audit process. It is the responsibility of the business leader and the business management. So, follow up with the companies should have been a communication from you to the business unit head, not from Don to the business and Don to you. But, we are going to work all of this out and in the meantime, we aren't going to get crosswise with each other.

Q: This passage is about how someone tries to settle the dispute in his company. _____ (T or F?)

3. Pete Weidler, a friend of mine, recommended that I contact you feeling that I may be able to help your group. Like you my background is not in e-commerce. I was most recently President of an oil company operating in Colombia. I am a manager in the classic sense, experienced in planning, project valuation, economic analysis, contract negotiations and acquisitions and divestitures. I am fluent in Spanish having grown up and worked in Latin America my whole life. I bring a keen awareness of the Latin culture and _____, a strong desire to continue working there. Do do this I'm considering a career change and can imagine no better company than yours to make that change smoothly.

Q: This passage is about someone's interest in studying Spanish. _____ (T or F?)

(1) That is (2) First of all (3) What's more (4) In addition (5) To begin with (6) What I'm saying is (7) Despite this/that (8) As a result (9) Because of that (10) But then again

4. However, it remains true to say that there is only one scenic mosaic in the east before c. 170- that from Eccles (no. 4; although this too could be later: appendix E, section 1.2 p.126). In the west of the province, there are no centrally placed and dominant single figures, or figured scenes. Instead, we encounter an apparently different concept of figural decoration. There are aquatic arrangements (probably influenced by the black and white mosaics, showing marine scenes, which were so popular in Italy, in the second and early-third century) and other such scenes e.g. the mask of Neptune, fish and dolphins of no. 18. _____, there are also examples of a more eclectic but "static" figural sequence, whose composite arrangements are extremely rich in figures but whose figures are linked only conceptually or are envisaged as nothing more than numerous, individual representations.

Q: This passage is about different ways of decoration. ____ (T or F?)

5. ...The important thing is to achieve the right balance. Commission Management commission rates vary, but a current average is 20 per cent of the artist's earnings. The agreement must specify that this 20 per cent should be on money actually received by the artist, not on items such as bad debts, and that there should be allowances for exactly what the commission is chargeable on in certain circumstances. For example, most managers seek commission on gross earnings, whereas artists' representatives attempt to reduce the manager's commission to net earnings from the band's live touring. Currently, touring is an extremely expensive business. _____, solicitors argue that managers should not take commission on what the artist is being paid each night, but on the money the band earns as profit after having paid for the PA, lights, accommodation and other expenses.

Q: This passage is about the commission rates for artists' managers. ____ (T or F?)

6. ...In some ways, although this mirrors the current situation for ordinary television, it is surprising that while global standards are an intense concern in multimedia development, the same is not true for HDTV. One of the major divisions has been between the Japanese and the Europeans. As long ago as May 1989, at the International Radio Consultative Committee meeting in Dusseldorf, Germany, European manufacturers blocked a Japanese move to get their 1125-line MUSE system accepted as a world standard. _____, MUSE has now been widely adopted in Japan and TV sets are being designed, tubes are being manufactured and large-screen projection systems are coming to market. A massive development programme is under way to produce the first large, LCD HDTV screens.

Q: This passage is about the standardization of HDTV. ____ (T or F?)

(1) As a result (2) What I'm saying is (3) To begin with (4) First of all (5) Because of that (6) In addition (7) That is (8) What's more (9) But then again (10) Despite this/that

7. ...whatever babe - i am concerned but not to the point of worrying my brains out - like you. i tend to trust the doctor in matters such as these- like when i had chest pains, the doctor ran tests to rule out significant problems and did notice an abnormality on one test. instead of running out and panicking i waited and took the tests again and they were normal. now i am not advocated waiting for the next test if that makes you uncomfortable - you should do what you want to do. _____, YOU ARE NOT GOING TO DIE !!! so stop worrying and let the tests confirm or deny any problem....

Q: This passage is about how someone is worried about his physical condition. ____ (T or F?)

8. ...Hey there. How are you today? I'm soooo tired but okay otherwise. I looked at my schedule for next week last night only to realize next week is going to be the most horrible week EVER!!!! I have a Greek and Roman History test Monday, a stats test Wednesday, and both a b. law test and an accounting test on Thursday. Way too many tests in the same week especially with two on the same day!!!!!!!! I'm sure I'll live through it but next week won't be fun for sure. By the way, I just thought I'd let you know that Joe Saporito left Lynne and the girls and is getting a divorce from Lynne. _____, Lynne, Lauren, and Rachel are going to come have Thanksgiving with Mom and I. I just thought you might want to know that since I know you like the girls a lot.

Q: This passage is about a business email between two companies. ____ (T or F?)

9. ...i am going to be really glad actually when school starts so that i can begin to get acclimated and don't feel like such a fish out of water. need to finalize when you're thinking about coming over.....i know it's sometime in oct- and kates said she has to leave on the 12th for a wedding thing. as i said, jeremy's wedding is the weekend of the 14th, and i know that was a week you discussed coming. i don't have tickets, but if it's physically poss for me to find a good ticket home, i prob. will have to- gogo started crying when i said i prob. couldn't come, and i promised i'd try. _____, i don't want to mess up the vacation plans you've been setting up for a long time. hopefully we can find a 'mutually beneficial' time for the visit- let me know what you're thinking, and we'll work from there!

Q: This passage is about someone's plan conflicts with his friend's. ____ (T or F?)

(1) Because of that (2) Despite this/that (3) In addition (4) But then again (5) What's more (6) To begin with (7) First of all (8) That is (9) As a result (10) What I'm saying is

10. ...The Queen had taken little or no interest in the appointment when the informal suggestion of Admiral Bryson came from Downing Street in 1989. When she heard the grumbles of the county landowners after the announcement, she may have wished she had done. Their apparent wealth notwithstanding, there are rumblings of unease within parts of the lieutenancy about the cost of it all. Except for an entitlement to a police car and driver when attending official functions, they receive no formal allowances. Most expect to meet the out-of-pocket expenses of the job themselves, but some feel the whole business is getting too costly. _____, there is the expense of the uniform, which London military tailors run up for about £2,000.

Q: This passage is about how landowners are struggling with their expense.

_____(T or F?)